Chapter 17

Rings and homomorphisms

Introduction

As you saw in Book C, Chapter 11, a ring can be thought of as a collection of objects that we can add, subtract and multiply, but not necessarily divide. There are many mathematical structures where this is precisely what we can do. The simplest and most familiar of these structures arises from the integers, where dividing one integer by another does not always give a third integer. From a mathematical point of view, the principal benefit we get from this constraint is the rich theory of prime numbers and the factorisation of integers in an essentially unique way into a product of primes.

Other mathematical structures where division is not closed include the rings of polynomials in one or several variables. Here the coefficients of the polynomials come themselves from a ring. They might, for example, be required to be integers, or they might belong to the field of real numbers. Polynomials are important in mathematics, and an interesting question is to what extent a ring of polynomials resembles the familiar ring of integers.

Of course, not being able to divide can be irksome, and in the case of the integers we can construct the smallest collection of objects that allows us to divide any integer by any (non-zero) integer. Such a quotient will not generally be an integer: it will be what we call a *rational number*. The collection of all the numbers obtained in this way forms the field of rational numbers. These numbers can be added, subtracted, multiplied and divided.

You may be familiar with the stories we tell children when introducing them to the rational numbers, in which they share out slices of pizzas and cakes, and which can leave them unsure why dividing by $\frac{1}{2}$, for example, has the same effect as multiplying by 2. Here we present a mathematical construction of the rationals that explains this informal notion of division much more systematically. Moreover, it can be applied to any integral domain, yielding a field that contains the integral domain and all the quotients of the elements of the integral domain. This field is called the *field of fractions* of the integral domain.

When dealing with polynomial rings where the coefficient ring is arbitrary, it is natural to wonder if the polynomials behave more or less as they do when the coefficients are integers. For example, can we factorise a polynomial as a product of irreducible polynomials, and can we do so in an essentially unique way? If the answer is 'yes', then this is one way in which the polynomial ring resembles the ring of integers. In fact, the answer depends on the ring of coefficients: if the ring of coefficient R is a unique factorisation domain then, as you will see, the polynomial ring $R[x]$ is also a unique factorisation domain. The proof is interesting because it makes use of F, the field of fractions of R, and allows us to compare the irreducibility of a polynomial in $R[x]$ with the same polynomial as an element of $F[x]$.

After working with polynomial rings, we consider rings more generally. Rings are, in particular, groups with a binary operation $+$, and we can look for objects and theorems in our theory of rings that resemble those that we find in group theory. A key notion is that of an *ideal*. In ring theory, ideals play a similar role to normal subgroups in group theory. As for groups, the ring of integers is a good source of examples: every non-trivial ideal in this ring consists of the multiples of a single positive integer. For example, the set of all multiples of 6 is an ideal in $\mathbb{Z}$. The notion of ideal is closely related to that of divisibility: for example, the ideal of all multiples of 3 in $\mathbb{Z}$ contains the ideal of all multiples of 6 in $\mathbb{Z}$, because every multiple of 6 is also a multiple of 3. In rings in which every ideal consists of multiples of a single element, the so-called *principal ideal rings*, there is a good theory of highest common divisors and least common multiples.

Like subgroups of a group, ideals of a ring can be used to construct other ideals. For example, we can form sums and products of ideals. Moreover, ideals arise as the kernels of ring homomorphisms, and there are analogues of the key isomorphism theorems in group theory. But perhaps the most important fact about ideals is that they give rise to new rings as quotients of a ring by one of its ideals, in the same way as normal subgroups give rise to quotient groups in group theory. In the example of the ideal of all multiples of 6 in $\mathbb{Z}$, the quotient ring is the ring of integers modulo 6.

An even richer example is a certain construction of the complex numbers from $\mathbb{R}$, the ring of real numbers. In this construction, we first build the polynomial ring $\mathbb{R}[x]$, then we construct the ideal I of all multiples of the polynomial $x^2 + 1$ by elements of $\mathbb{R}[x]$. Finally, we quotient $\mathbb{R}[x]$ by I. As you will see in Section 4.4, the resulting ring is isomorphic to the ring of complex numbers.

There are two properties of a prime number that you met in the wider context of rings in Book C, Chapter 12, Section 2.2. The first is irreducibility: a number is irreducible if it cannot be factored as a product in a non-trivial way. The second is primality: a number p is prime if whenever it divides a product ab it divides one of the factors – either p divides a or p divides b. As you saw in Book C, Chapter 13, in a general ring it can happen that an element is irreducible but does not have the second property. The way to pursue the analogy with the integers is to introduce the notion of *prime* ideal, so the chapter concludes with a study of prime ideals.

Chapter guide

Although Section 4 is substantially longer than the other sections, much of the material is very closely related to the material on quotient groups that was covered in Book B. Section 4 requires about twice as much study time as Section 1. Sections 2 and 3 together require about the same amount of study time as Section 1.

1 Fields of fractions

In this section we will introduce the field of fractions of an integral domain. This field is obtained by expanding the integral domain to a larger ring in which all non-zero elements have multiplicative inverses. For example, consider the relationship between the ring $\mathbb{Z}$ of integers and the field $\mathbb{Q}$ of rational numbers. We know that $\mathbb{Z} \subset \mathbb{Q}$, and that all non-zero elements in $\mathbb{Q}$ have multiplicative inverses. The inclusion $\mathbb{Z} \subset \mathbb{Q}$ is often called an *embedding*, and we say that the ring of integers is *embedded* into the field of rational numbers. The elements of $\mathbb{Q}$, that is, the rational numbers, are fractions formed by the integers. This way of looking at $\mathbb{Z}$ and $\mathbb{Q}$ leads to a far-reaching generalisation, which is the construction of a field of fractions from an integral domain. This construction has an important application to polynomial rings over unique factorisation domains.

1.1 The field of fractions of an integral domain

Let us start by recalling how the rational numbers are constructed from the integers. We know that, for example, $\frac{2}{3}$ and $\frac{6}{9}$ represent the same rational number. More generally, two fractions $\frac{a}{b}$ and $\frac{c}{d}$ represent the same rational number if, and only if, $ad = bc$. So a rational number can be thought of as an infinite set of fractions $\frac{a}{b}$, each of which represents the number.

In mathematically rigorous terms, what we do is introduce an equivalence relation on the set of ordered pairs

$$\{(a, b) : a, b \in \mathbb{Z}, b \neq 0\}$$

by stipulating that two pairs (a, b) and (c, d) are equivalent if $ad = bc$. A rational number is defined as an equivalence class under this relation. We then introduce addition and multiplication of rational numbers in the usual way and prove that the two operations are well defined, that is, they do not depend on the choice of ordered pairs representing the rational numbers.

In what follows, we mimic this construction of the rational numbers from the integers, but instead of the integers we start off with a general integral domain.

Recall from Book C, Chapter 12, Definition 2.8 that $(D, +, \cdot)$ is an integral domain if it is a commutative ring with unity and with no zero divisors. As in Book C, throughout this book we usually simplify notation by using juxtaposition rather than $\cdot$ for the second ring operation.

Now let S_D be the set of all ordered pairs (a, b) of elements of D such that $b \neq 0$, that is

$$S_D = \{(a, b) : a, b \in D,\ b \neq 0\}.$$

We introduce a binary relation $\sim$ on the set S_D as follows:

$$(a, b) \sim (c, d) \text{ if, and only if, } ad = bc.$$

Our aim is to show that $\sim$ is an equivalence relation, namely that it is reflexive, symmetric and transitive.

Exercise 1.1

Let $(D, +, \cdot)$ be an integral domain and let

$$S_D = \{(a, b) : a, b \in D,\ b \neq 0\}.$$

Show that the relation $\sim$ on the set S_D defined by

$$(a, b) \sim (c, d) \text{ if, and only if, } ad = bc$$

is reflexive and symmetric.

Proposition 1.1

Let $(D, +, \cdot)$ be an integral domain and let

$$S_D = \{(a, b) : a, b \in D,\ b \neq 0\}.$$

Then the relation $\sim$ on the set S_D defined by

$$(a, b) \sim (c, d) \text{ if, and only if, } ad = bc$$

is an equivalence relation.

Proof We proved that the relation is symmetric and reflexive in the solution to Exercise 1.1, so we just need to show transitivity. Let (a, b), (c, d) and (e, f) be elements of S_D such that $(a, b) \sim (c, d)$ and $(c, d) \sim (e, f)$. To prove that $\sim$ is transitive we need to show that $(a, b) \sim (e, f)$.

Since $(a, b) \sim (c, d)$, we have that $ad = bc$. Multiplying both sides of this equality on the right by f gives

$$adf = bcf.$$

Similarly, $(c, d) \sim (e, f)$ gives $cf = de$, and left multiplication by b gives

$$bcf = bde.$$

Therefore $adf = bde$, and so $afd = bed$ since multiplication is commutative in D. Since D is an integral domain, $af = be$ by cancellation (Book C, Chapter 12, Lemma 2.11). Thus $(a, b) \sim (e, f)$ as required. ■

We now know that $\sim$ is an equivalence relation, and so we let F_D denote the set of equivalence classes on S_D induced by $\sim$, that is,

$$F_D = S_D/\sim .$$

For any $(a, b) \in S_D$, the equivalence class in F_D containing the element (a, b) is denoted by $[a, b]$:

$$[a, b] = \{(c, d) \in S_D :\ (a, b) \sim (c, d)\}.$$

On this new set F_D we introduce the operations of addition $+$ and multiplication $\cdot$ as follows:

$$[a,b]+[c,d]=[ad+bc,bd]$$
$$[a,b]\cdot[c,d]=[ac,bd].$$

We will again simplify notation by using juxtaposition in place of $\cdot$ for multiplication.

Our next aim is to show that these two operations are well defined, that is, independent of the choice of representatives for equivalence classes.

As regards addition, we have to show that if $[a,b]=[a_1,b_1]$ and $[c,d]=[c_1,d_1]$, then $[ad+bc,bd]=[a_1d_1+b_1c_1,b_1d_1]$.

Our assumptions say that $(a,b)\sim(a_1,b_1)$ and $(c,d)\sim(c_1,d_1)$, which implies that

$$ab_1=ba_1 \text{ and } cd_1=c_1d \qquad (*)$$

and we would like to conclude that $(ad+bc,bd)\sim(a_1d_1+b_1c_1,b_1d_1)$. That is, we would like to prove

$$(ad+bc)(b_1d_1)=(bd)(a_1d_1+b_1c_1).$$

Since D is an integral domain, $bd\neq 0$, so both operations are closed in S_D.

Now

$$\begin{aligned}(ad+bc)(b_1d_1)&=adb_1d_1+bcb_1d_1\\&=ab_1dd_1+bb_1cd_1\\&=ba_1dd_1+bb_1c_1d \quad \text{by } (*)\\&=bda_1d_1+bdb_1c_1\\&=(bd)(a_1d_1+b_1c_1),\end{aligned}$$

as required.

The next exercise asks you to check that multiplication is well defined.

Exercise 1.2

Using the notation just given, show that $[a,b][c,d]=[ac,bd]$ is well defined, that is, it is independent of the choice of equivalence classes.

We now want to show that the algebraic structure $(F_D,+,\cdot)$ is a field, that is, a commutative ring with unity where every non-zero element has a multiplicative inverse (see Book C, Chapter 11, Definition 1.14). In order to do this, we start by showing that cancellation holds with these operations, and we describe the equivalence class of $(0,1)$, which will be the additive identity.

Lemma 1.2

Let D be an integral domain and let $S_D = \{(a, b) : a, b \in D,\ b \neq 0\}$. Let $F_D = \{[a, b] : (a, b) \in S_D\}$ denote the set of equivalence classes on S_D induced by the relation $\sim$ defined by

$$(a, b) \sim (c, d) \text{ if, and only if, } ad = bc.$$

(a) If $a, b, c \in D$ with $b \neq 0$ and $c \neq 0$, then $(a, b) \sim (ac, bc)$.

(b) If $a, b, d \in D$ with $b \neq 0$ and $d \neq 0$, then $(a, b) \sim (0, d)$ if, and only if, $a = 0$.

Proof

(a) Since D is commutative, $abc = bac$ for all a, b and c in D. This implies that $(a, b) \sim (ac, bc)$.

(b) We have $(a, b) \sim (0, d)$ if, and only if, $ad = b0 = 0$. Since D has no zero divisors and $d \neq 0$, $ad = 0$ if, and only if, $a = 0$.

Therefore $(a, b) \sim (0, d)$ if, and only if, $a = 0$, as required. ■

Proposition 1.3

Let D be an integral domain and let S_D and F_D be as in Lemma 1.2. Let $+$ and $\cdot$ be binary operations on F_D defined by

$$[a, b] + [c, d] = [ad + bc, bd]$$
$$[a, b] \cdot [c, d] = [ac, bd].$$

Then F_D is a field with the property that every element can be written in the form

$$[a, 1][b, 1]^{-1}$$

for some $a, b \in D$, $b \neq 0$.

Proof We know from the solution to Exercise 1.2 and the discussion preceding it that the two binary operations $+$ and $\cdot$ are well defined. We now check the axioms for a field given in Definitions 1.1 and 1.14 of Book C, Chapter 11.

Axioms for addition

R1 Closure For all $[a,b],[c,d] \in F_D$, we have that $bd \neq 0$ because $b \neq 0$ and $d \neq 0$, and D has no zero divisors. Therefore

$$[a,b]+[c,d] = [ad+bc, bd] \in F_D.$$

R2 Associativity For all $[a,b],[c,d],[e,f] \in F_D$,

$$\begin{aligned}[a,b]+([c,d]+[e,f]) &= [a,b]+[cf+de, df] \\ &= [adf+b(cf+de), bdf] \\ &= [adf+bcf+bde, bdf] \\ &= [(ad+bc)f+bde, bdf] \\ &= [ad+bc, bd]+[e,f] \\ &= ([a,b]+[c,d])+[e,f].\end{aligned}$$

Note that this argument uses both associativity and distributivity in D.

R3 Additive identity For all $[a,b] \in F_D$,

$$[a,b]+[0,1] = [a,b] = [0,1]+[a,b].$$

R4 Additive inverses For each $[a,b] \in F_D$,

$$[a,b]+[-a,b] = [0,1] = [-a,b]+[a,b].$$

R5 Commutativity For all $[a,b],[c,d] \in F_D$,

$$\begin{aligned}[a,b]+[c,d] &= [ad+bc, bd] \\ &= [cb+da, db] \text{ from the commutativity of } D \\ &= [c,d]+[a,b].\end{aligned}$$

Axioms for multiplication

R6 Closure For all $[a,b],[c,d] \in F_D$,

$$[a,b][c,d] = [ac, bd] \in F_D \text{ since } b \neq 0 \text{ and } d \neq 0.$$

As in the proof of additive closure, here we are using the fact that D has no zero divisors, which is true because D is an integral domain.

R7 Associativity For all $[a,b],[c,d],[e,f] \in F_D$,

$$\begin{aligned}[a,b]([c,d][e,f]) &= [a,b][ce, df] \\ &= [ace, bdf] \\ &= [ac, bd][e,f] \\ &= ([a,b][c,d])[e,f].\end{aligned}$$

This argument uses associativity in D.

R8 Multiplicative identity For all $[a,b] \in F_D$,

$$[a,b][1,1] = [a,b] = [1,1][a,b].$$

Axioms combining addition and multiplication

R9 Distributive laws For all $[a,b],[c,d],[e,f] \in F_D$,

$$\begin{aligned}
[a,b]([c,d]+[e,f]) &= [a,b][cf+de,df] \\
&= [acf+ade,bdf]. \\
([a,b][c,d])+([a,b][e,f]) &= [ac,bd]+[ae,bf] \\
&= [acbf+bdae,bdbf] \\
&= [acf+dae,bdf] \text{ using Lemma 1.2(a)} \\
&= [acf+ade,bdf] \text{ by the commutativity of } D.
\end{aligned}$$

This shows that the left distributive law holds. The right distributive law follows from left distributivity by the commutativity of D.

We are now left to check axioms R10 and R11.

R10 Commutativity For all $[a,b],[c,d] \in F_D$,

$$[a,b][c,d] = [ac,bd] = [ca,db] = [c,d][a,b] \text{ by the commutativity of } D.$$

R11 Multiplicative inverses Let $[a,b]$ be a non-zero element of F_D. Then Lemma 1.2(b) implies that $a \neq 0$, so $[b,a] \in F_D$. Moreover,

$$\begin{aligned}
[a,b][b,a] &= [ab,ba] = [1,1] \quad \text{and} \\
[b,a][a,b] &= [ba,ab] = [1,1]
\end{aligned}$$

by Lemma 1.2(a). Thus $[b,a]$ is the multiplicative inverse of $[a,b]$, that is $[b,a] = [a,b]^{-1}$.

Thus, all the field axioms hold in F_D. Now let $[a,b] \in F_D$. Then

$$[a,b] = [a,1][1,b] = [a,1][b,1]^{-1},$$

and so every element of F_D has the required form. ■

We are used to thinking of the integers as a subring of the rationals by identifying the integer a with the rational $\frac{a}{1}$. In terms of the equivalence relation that we defined on $\{(a,b) : a,b \in \mathbb{Z}, b \neq 0\}$, we identify the elements of $\mathbb{Z}$ with the elements of the set $\{[a,1] : a \in \mathbb{Z}\}$.

Similarly, in the general case the set

$$\{[a,1] \in F_D : a \in D\}$$

is a copy of D in F_D, and, just as with rational numbers, we identify the elements of this set with the elements of D. Now, we know from Proposition 1.3 that every element $[a,b]$ of F_D can be written as $[a,1][b,1]^{-1}$. So we can extend our identification to the whole of F_D by thinking of $[a,b] \in F_D$ as ab^{-1}, where a and $b \in D$ and $b \neq 0$.

This identification is the defining feature of a field of fractions, and it can be expressed in rigorous mathematical terms using the notion of a *ring isomorphism*.

Definition 1.4 *Ring isomorphism*

Let R and S be rings. Then we say that R and S are **isomorphic** if there is a bijection $\theta : R \to S$ such that, for all $a, b \in R$,

- $\theta(a + b) = \theta(a) + \theta(b)$
- $\theta(ab) = \theta(a)\theta(b)$
- $\theta(1_R) = 1_S$.

The bijection θ is said to be a **ring isomorphism**.

We often refer to the properties of addition and multiplication in Definition 1.4 as the **homomorphism** or **morphism properties**.

With this definition we can now check that for a given integral domain D the field of fractions F_D is unique up to isomorphism.

Proposition 1.5

Let E be a field, and let D be a subring of E such that

1. D is an integral domain
2. every element of E can be written as ab^{-1} with $a, b \in D$.

Then E is isomorphic to F_D.

Proof We show that the map

$$\begin{aligned} \theta : \quad E &\to F_D \\ ab^{-1} &\mapsto [a, b] \end{aligned}$$

is an isomorphism. Let ab^{-1} and $cd^{-1} \in E$. Then

$$\begin{aligned} ab^{-1} + cd^{-1} &= add^{-1}b^{-1} + cbb^{-1}d^{-1} \\ &= ad(bd)^{-1} + bc(bd)^{-1} \\ &= (ad + bc)(bd)^{-1}. \end{aligned}$$

This calculation makes use of the associative, commutative and distributive laws in E. We use this equality to check that θ satisfies the properties in Definition 1.4. Remember that $[1, 1]$ is the multiplicative identity of F_D.

- $$\begin{aligned} \theta(ab^{-1} + cd^{-1}) &= \theta((ad + bc)(bd)^{-1}) \\ &= [ad + bc, bd] \\ &= [a, b] + [c, d] \end{aligned}$$
- $$\begin{aligned} \theta(ab^{-1}cd^{-1}) &= \theta(ac(bd)^{-1}) \\ &= [ac, bd] \\ &= [a, b][c, d] \end{aligned}$$
- $$\begin{aligned} \theta(1) &= \theta(1 \cdot 1^{-1}) \\ &= [1, 1] \end{aligned}$$

We have to show that θ is a bijection. Since any $[a, b] \in F_D$ is the image of $ab^{-1} \in E$, θ is onto.

Moreover, $\theta(ab^{-1}) = \theta(cd^{-1}) \implies [a, b] = [c, d]$
$\implies ad = bc$
$\implies ab^{-1} = cd^{-1}$,
which shows that θ is one–one.

Thus θ is a bijection, and so it is a ring isomorphism. ■

Proposition 1.5 ensures that an integral domain has a unique field of fractions up to isomorphism, so the following definition is justified.

Definition 1.6 *Field of fractions*

Let D be an integral domain. Then a field F_D is called the **field of fractions** of D if D is isomorphic to a subring A of F_D and every element of F_D can be written as ab^{-1} where $a, b \in A$ and $b \neq 0$.

The isomorphism whose existence we proved in Proposition 1.5 allows us to use the definite article 'the' before 'field of fractions'.

In the preceding discussion we have used the notation $[a, b]$ for the equivalence class corresponding to an element of F_D. It is, however, usual to adopt the fraction notation $\frac{a}{b}$, and we will do this from now on.

We will also stop using the notation F_D and simply refer to F as the field of fractions of an integral domain D.

Example 1.7 *Two fields of fractions*

(a) Let $\mathbb{Z}$ denote the ring of integers. The field of rational numbers $\mathbb{Q}$ is isomorphic to the field of fractions F of the integers.

(b) The field of *rational functions*, which has underlying set

$$\left\{ \frac{f(x)}{g(x)} : f(x), g(x) \in \mathbb{Q}[x], g(x) \neq 0 \right\},$$

is the field of fractions of $\mathbb{Q}[x]$. It is often denoted by $\mathbb{Q}(x)$.

Worked Exercise 1.8

Worked Exercise 1.9 of Book C, Chapter 11 introduced the ring of Gaussian integers

$$\mathbb{Z}[i] = \{a + ib : a, b \in \mathbb{Z}\}.$$

What is the field of fractions of $\mathbb{Z}[i]$?

Solution

By the remark immediately preceding Example 1.7, we may identify the elements of the field of fractions of the Gaussian integers with fractions of the form

$$\frac{a+bi}{c+di},$$

where $a, b, c, d \in \mathbb{Z}$ and c, d are not both equal to 0, so that $c + id \neq 0$.

If we consider the fractions with $d = 0$, we see that the field of fractions contains all the elements of the form $a_1 + b_1 i$, where $a_1, b_1 \in \mathbb{Q}$. On the other hand, multiplying the numerator and the denominator of $\frac{a+bi}{c+di}$ by $c - di$ gives

$$\frac{(a+bi)(c-di)}{(c+di)(c-di)} = \frac{ac+bd+i(bc-ad)}{c^2+d^2} = \frac{ac+bd}{c^2+d^2} + i\left(\frac{bc-ad}{c^2+d^2}\right).$$

This shows that these fractions always have the form $a_1 + b_1 i$ where $a_1, b_1 \in \mathbb{Q}$.

We conclude that the field of fractions of the Gaussian integers $\mathbb{Z}[i]$ is isomorphic to

$$\mathbb{Q}[i] = \{a + bi : a, b \in \mathbb{Q}\},$$

which is known as the **field of Gaussian numbers**.

Alternatively, we can show that $\mathbb{Q}[i]$ is the field of fractions for $\mathbb{Z}[i]$ by proving that $\mathbb{Q}[i]$ is a field and that every element can be written in the form ab^{-1} where $a, b \in \mathbb{Z}[i]$.

We leave you to check that $\mathbb{Q}[i]$ is an integral domain. Then, in order to show that $\mathbb{Q}[i]$ is a field, we just need to show that every non-zero element has a multiplicative inverse.

Let $a + bi \in \mathbb{Q}[i]$ be such that $a + bi \neq 0$, so that $a^2 + b^2 \neq 0$. Then

$$(a+bi)\left(\frac{a}{a^2+b^2} - \frac{b}{a^2+b^2}i\right) = \frac{(a+ib)(a-ib)}{a^2+b^2} = 1.$$

Thus we have shown that every non-zero element $a + bi$ of $\mathbb{Q}[i]$ has a multiplicative inverse in $\mathbb{Q}[i]$. Therefore $\mathbb{Q}[i]$ is a field.

Now let $a + bi \in \mathbb{Q}[i]$ with $a = \frac{a_1}{a_2}$, $b = \frac{b_1}{b_2}$ where $a_1, a_2, b_1, b_2 \in \mathbb{Z}$ and $a_2 \neq 0$, $b_2 \neq 0$. Then

$$a + bi = \frac{a_1}{a_2} + \frac{b_1}{b_2}\, i = \frac{a_1 b_2 + b_1 a_2 i}{a_2 b_2} = (a_1 b_2 + b_1 a_2 i)(a_2 b_2)^{-1}.$$

Thus we can write every element of $\mathbb{Q}[i]$ in the form cd^{-1} where $c, d \in \mathbb{Z}[i]$. Hence $\mathbb{Q}[i]$ is the field of fractions of $\mathbb{Z}[i]$ by Definition 1.6.

Exercise 1.3

(a) Show that there exist $f \in \mathbb{Q}$, and $p(x) \in \mathbb{Z}[x]$ such that

$$\frac{1}{2} + \frac{2}{3}x + \frac{1}{5}x^3 = fp(x).$$

(b) Let $R[x]$ be a polynomial ring over an integral domain R and let F be the field of fractions of R. Show every element $q(x) \in F[x]$ can be written as $fp(x)$ where $f \in F$ and $p(x) \in R[x]$.

Using the fraction notation might make you wonder whether we can always write the elements of a field of fractions as fractions in their 'lowest terms', in some sense of 'lowest'. In the next exercise you are asked to show that this can be done when D is a unique factorisation domain, or UFD – recall Definition 4.1 of Book C, Chapter 12.

Exercise 1.4

Let D be a UFD and let F be the field of fractions of D. Let $\frac{a}{b} \in F$ be a non-zero element, so that $a \neq 0$. Show that there exist $c, d \in D$ such that

$$\frac{a}{b} = \frac{c}{d}$$

and $\text{hcf}(c, d) = u$, where u is a unit in D.

When two elements c and d of a unique factorisation domain D are such that $\text{hcf}(c, d)$ is a unit, we say that $\frac{c}{d} \in F_D$ is a **fraction in its lowest terms**.

Exercise 1.5

Let D be a UFD and let F be the field of fractions of D. Let S be a finite set of fractions,

$$S = \left\{ \frac{a_i}{b_i} : a_i, b_i \in D,\ b_i \neq 0,\ i = 1, \ldots, n \right\}.$$

(a) Show that there is $b \in D$ such that

(i) $b = b_i c_i$ for some c_i, $i = 1, \ldots, n$;

(ii) if $d = b_i d_i$ for some d_i, $i = 1, \ldots, n$ then $d = bc$ for some $c \in D$.

That is, show that $\text{lcm}(b_1, b_2, \ldots, b_n)$ exists.

(b) Let $\frac{a_i}{b_i} \in S$. Show that $\frac{a_i}{b_i} = \frac{a_i c_i}{b}$.

Hint: for part (a), use the result in Exercise 4.1 in Book C, Chapter 12.

The element $b \in D$ in Exercise 1.5 is known as a **lowest common denominator** of the set of fractions S.

1.2 Polynomial rings over a UFD

In Theorem 4.11 of Book C, Chapter 12, we proved that $\mathbb{Z}[x]$ is a UFD. A key step in the proof is to consider a polynomial in $\mathbb{Z}[x]$ as a polynomial in $\mathbb{Q}[x]$ in order to use the fact that $\mathbb{Q}[x]$ is a Euclidean domain. Any UFD is an integral domain and so it has a field of fractions. This means we can generalise this key step to any UFD to prove the main theorem of this subsection, namely that if R is a UFD then $R[x]$ is a UFD. We start by generalising some definitions and Gauss's Lemma (Theorem 4.13 of Book C, Chapter 11), which is the key preliminary result linking factorisation in $\mathbb{Z}[x]$ with factorisation in $\mathbb{Q}[x]$.

Definition 1.9 *Primitive polynomial*

Let R be a unique factorisation domain, and let

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$

be a polynomial in $R[x]$.

(a) The **content** of f is a highest common factor of all of its coefficients in R:

$$c_f = \text{hcf}(a_0, a_1, \ldots, a_n).$$

(b) $f(x)$ is **primitive** if $c_f = u$ for some unit in $u \in R$. In other words, $f(x)$ is primitive if the only common factors of its coefficients are units of R.

In general, highest common factors in a UFD are only unique up to multiplication by a unit, so the same is true of the content. This is why we define the content as 'a' highest common factor of the coefficients (rather than 'the' highest common factor).

As a consequence of Definition 1.9, if R is a unique factorisation domain and $f(x) \in R[x]$, we can write $f(x) = c_f g(x)$ where $g(x)$ is a primitive polynomial in $R[x]$.

The following lemma is the natural extension to UFDs of Lemma 4.12 of Book C, Chapter 11.

Lemma 1.10

Let R be a unique factorisation domain. Then f and g are primitive polynomials in $R[x]$ if, and only if, their product fg is primitive.

Proof The proof is the same as that of Lemma 4.12 of Book C, Chapter 11 with $\mathbb{Z}$ replaced by R and with $p \in \mathbb{Z}$ replaced by an irreducible element $r \in R$. ■

Lemma 1.10 enables us to prove a generalised version of Gauss's Lemma. The proof is very similar to that of Theorem 4.13 in Chapter 11, but we will repeat it here, as the general case uses the field of fractions of a UFD.

Lemma 1.11 *Gauss's Lemma for UFDs*

Let R be a UFD and let F be the field of fractions of R. Let $f(x)$ be a primitive polynomial in $R[x]$. If we can write $f = gh$ with $g, h \in F[x]$, then we can write $f = GH$ with $G, H \in R[x], \deg(g) = \deg(G)$ and $\deg(h) = \deg(H)$.

Proof Since $g(x)$ is a polynomial in $F[x]$, by the solution to Exercise 1.5 we can find a lowest common denominator, b say, of all its coefficients. Then we can write $g(x) = \frac{a}{b}\hat{G}(x)$, where $\hat{G}(x)$ is a primitive polynomial and we may assume that $\frac{a}{b}$ is a fraction in its lowest terms by the solution to Exercise 1.4.

Similarly we can write $h(x) = \frac{c}{d}\hat{H}(x)$ where $\hat{H}(x)$ is a primitive polynomial in $R[x]$ and $\frac{c}{d}$ is a fraction in lowest terms. Note that $\deg(g) = \deg(\hat{G})$ and $\deg(h) = \deg(\hat{H})$. Hence

$$f(x) = g(x)h(x) = \frac{ac}{bd}\hat{G}(x)\hat{H}(x).$$

We now prove that $f = u\hat{G}\hat{H}$ for some unit $u \in R$, and so we need to show that $ac = ubd$. Rearranging, we find that $bdf = ac\hat{G}\hat{H}$. Since f is primitive, the content of $bdf(x)$ is bd. Moreover, by Lemma 1.10, $\hat{G}\hat{H}$ is primitive, and so the content of $ac\hat{G}(x)\hat{H}(x)$ is ac. Since the polynomials $bdf(x)$ and $ac\hat{G}(x)\hat{H}(x)$ are equal, so must their contents be equal, and hence $ac = ubd$ for some unit $u \in R$ as required.

Taking $G = u\hat{G}$ and $H = \hat{H}$ gives the required factorisation $f = GH$. ■

We are almost ready to prove that a polynomial ring over a UFD is itself a UFD. The proof that $\mathbb{Z}[x]$ is a UFD given in Book C, Chapter 12, Subsection 4.2 used the following steps:

1. Every polynomial in $\mathbb{Z}[x]$ has a factorisation into irreducible polynomials in $\mathbb{Z}[x]$ (Lemma 4.8).
2. Any irreducible polynomial in $\mathbb{Z}[x]$ is prime in $\mathbb{Z}[x]$ (Proposition 4.9). This is proved in two stages:
   (a) Let $p \in \mathbb{Z}$. If p is prime in $\mathbb{Z}$, then it is prime in $\mathbb{Z}[x]$ (Lemma 4.10).
   (b) Let f be an irreducible polynomial in $\mathbb{Z}[x]$ with $\deg(f) \geq 1$. Then f is prime in $\mathbb{Z}[x]$.

The proof in Chapter 12 also used Theorem 4.2, which we restate here for reference.

Theorem 1.12 *Theorem 4.2 of Chapter 12*

Let R be an integral domain in which every non-zero non-unit element has a factorisation into irreducible elements. Then R is a UFD if, and only if, every irreducible element in R is prime.

Our proof that a polynomial ring over a UFD is a UFD mimics that of Theorem 4.11. We state each of the above steps as a lemma.

Lemma 1.13

Let R be a UFD. Then every polynomial in $R[x]$ has a factorisation into irreducible polynomials in $R[x]$.

Proof Let $f(x) \in R[x]$ be a non-trivial element of $R[x]$. Then

$$f(x) = c_f f^*(x),$$

where $f^*(x)$ is a primitive polynomial. Since $c_f \in R$, either c_f is a unit, or it factorises uniquely (up to multiplication by a unit) in R as

$$c_f = p_1 p_2 \cdots p_s,$$

where each p_i is irreducible in R. Since p_i is irreducible in R, it is also irreducible in $R[x]$. We now have $f(x) = p_1 p_2 \cdots p_s f^*(x)$, so it suffices to show that every primitive polynomial in $R[x]$ has a factorisation into irreducibles. We will use a minimal counter-example argument. Let

$$S = \{f \in R[x] : f(x) \text{ is primitive, non-zero, not a unit and cannot be written as a product of irreducibles}\}.$$

We show that S is empty. Suppose for a contradiction that $f(x) \in S$. Then, by the Well-Ordering Principle (see Book A, Chapter 1, Subsection 3.2), we can pick f such that $\deg(f)$ is least among the degrees of polynomials in S.

Now, since f cannot be written as a product of irreducibles, it cannot itself be irreducible. Therefore we have a proper factorisation $f = ab$ for some polynomials $a, b \in R[x]$.

Note that we cannot have $\deg(a) = 0$, as then $a \in R$, and since f is primitive this would imply that a is a unit, so the factorisation $f = ab$ would not be proper. Thus $\deg(a) \geq 1$ and, symmetrically, $\deg(b) \geq 1$. Furthermore, by Lemma 1.10, both a and b must be primitive.

From $\deg(f) = \deg(ab) = \deg(a) + \deg(b)$ we have $\deg(a) < \deg(f)$ and $\deg(b) < \deg(f)$. Our choice of f implies that each of a and b is either irreducible or can be written as a product of irreducibles. This implies that $f = ab$ can be written as a product of irreducibles. This contradiction shows that S does not have an element of minimal degree, and therefore S must be empty, that is, all primitive polynomials can be written as a product of irreducibles. ■

Having established the existence of a factorisation into irreducibles for every element of $R[x]$, it remains to show that every irreducible element in $R[x]$ is prime in $R[x]$.

We start with a lemma that applies to polynomials of degree 0.

Lemma 1.14

Let R be an integral domain and let $R[x]$ be the ring of polynomials in one variable over R. If $p \in R$ is prime in R then it is prime in $R[x]$.

Proof We will use a minimal counter-example argument. Suppose that p is not prime in $R[x]$. Then there exist polynomials $f(x), g(x)$ with $\deg f = m$ and $\deg g = n$ such that p divides $f(x)g(x)$ but p divides neither $f(x)$ nor $g(x)$. Using the Well-Ordering Principle we can choose polynomials f and g so that $m + n$ is the least possible value for which this holds. Note that since p is prime in R, $m + n > 0$.

Now let $f(x) = a_m x^m + \cdots + a_0$ and
$g(x) = b_n x^n + \cdots + b_0$.
If p divides $f(x)g(x)$, then p must divide every coefficient in $f(x)g(x)$. Thus p divides the leading coefficient $a_m b_n$, and so, since p is prime in R, p divides either a_m or b_n. Without loss of generality, suppose that p divides a_m. Note that $m > 0$ since otherwise p divides $f(x)$. Then p divides

$$f(x)g(x) - a_m x^m g(x) = (f(x) - a_m x^m)g(x).$$

Now $f(x) - a_m x^m$ is of lower degree than $f(x)$. Since we chose $m + n$ to be minimal, and since p does not divide $g(x)$, we have that p divides $f(x) - a_m x^m$. However, this means that p divides

$$f(x) - a_m x^m + a_m x^m = f(x),$$

which contradicts our original assumption. Thus p must be prime in $R[x]$. ■

Lemma 1.15

Let R be a UFD, and f be an irreducible element in $R[x]$. Then f is prime in $R[x]$.

Proof Suppose f divides gh, where $g, h \in R[x]$. We need to show that f divides one of g or h. If $\deg(f) = 0$, then $f \in R$, and therefore f is prime in R by Theorem 4.2 of Book C, Chapter 12. Hence f is prime in $R[x]$ by Lemma 1.14.

Now suppose $\deg(f) > 0$. Then $f(x) = c_f g(x)$, where $g(x)$ is primitive. But f is irreducible, so c_f must be a unit and f itself is primitive in $R[x]$.

Let F denote the field of fractions of R. By Lemma 1.11, f is irreducible in $F[x]$, and since $F[x]$ is a Euclidean domain (see Book C, Chapter 12, Example 3.2(c)) f must be prime in $F[x]$ by Proposition 3.11 of Book C, Chapter 12. Thus since f divides gh in $F[x]$, either f divides g or f divides h, and without loss of generality we assume that f divides g.

Then there exists $s \in F[x]$ such that $fs = g$. We can write $s(x) = ct(x)$, where $c \in F$ and $t(x) \in R[x]$ is primitive. Therefore the following equation holds in $F[x]$:

$$f(x)(ct(x)) = g(x).$$

Since $c \in F$, we can write $c = \frac{a}{b}$ for some non-zero $a, b \in R$ with $\mathrm{hcf}(a, b) = 1$ (see Exercise 1.4). We want to show $c \in R$. Multiplying through by b gives

$$f(x)(at(x)) = bg(x).$$

We now consider the content of the polynomials on either side of this equation. On the left, the content is a, since f and t are primitive. On the right the content is bc_g, where c_g is the content of g, and this must be equal to the content on the left, so $a = bc_g$. However, $c_g \in R$, and so $c = \frac{a}{b} = c_g \in R$, and therefore $ct(x) \in R[x]$. Hence f divides g in $R[x]$, so f is prime in $R[x]$. ■

Theorem 1.16

Let R be a UFD. Then the polynomial ring $R[x]$ is a UFD.

Proof By Lemma 1.13, every non-zero, non-unit element of $R[x]$ has a factorisation into irreducibles. By Lemma 1.15, every irreducible element of $R[x]$ is prime. Therefore $R[x]$ is a UFD by Theorem 1.12. ■

We finish this section with an example of a ring, related to polynomial rings, that is also a UFD.

Example 1.17 *The ring of Laurent polynomials*

Let F be a field. Then

$$F[x, x^{-1}] = \{a_m x^m + \cdots + a_n x^n : a_i \in F \text{ for } i = m, \ldots, n \text{ with } m, n \in \mathbb{Z},\ m < n\}$$

is a ring, called the ring of **Laurent polynomials** over F. Note that since $m, n \in \mathbb{Z}$ the powers of x may be negative. For example,

$$5x^{-3} + 2x^{-1} + 5 + 7x^2$$

is a Laurent polynomial with $m = -3$ and $n = 2$.

$F[x, x^{-1}]$ is a subset of the field of fractions of $F[x]$, so to see that it is a ring we only need to observe that it is closed under both addition and multiplication, includes both the additive and multiplicative identities and includes additive inverses.

Laurent polynomials are named after the mathematician Pierre Alphonse Laurent (1813–1854). They are related to *Laurent series*, which are an important tool in complex analysis. A Laurent polynomial can be viewed as a Laurent series in which there are only finitely many non-zero coefficients.

We claim that $F[x, x^{-1}]$ is a Euclidean domain, and hence a UFD. To show that it is a Euclidean domain we need to show that it has a Euclidean norm (see Book C, Chapter 12, Definitions 2.14 and 3.1).

Let $f(x) = a_m x^m + \cdots + a_n x^n$ be in $F[x, x^{-1}]$ with a_m and a_n both non-zero, and define

$$N(f(x)) = n - m.$$

We claim that N is a Euclidean norm. Note that $N(x^v f(x)) = N(f(x))$ for all $v \in \mathbb{Z}$. The intuition is that the norm of a Laurent polynomial is, in some sense, a measure of its length, and multiplying the polynomial by a power of x just shifts it in one direction or the other without changing the length. Moreover, if $f(x) \in F[x]$, rather than $F[x, x^{-1}]$, and has a non-zero constant term, then $N(f(x)) = \deg f(x)$. More generally, if $f(x) \in F[x]$ then $N(f(x)) \leq \deg f(x)$.

To show N is a Euclidean norm, we need to show that, for $f, g \in F[x, x^{-1}]$,

(i) $N(f) \leq N(fg)$;

(ii) there exists $q, r \in F[x, x^{-1}]$ such that $f = gq + r$, and either $r = 0$ or $N(r) < N(g)$.

Let $f(x) = a_m x^m + \cdots + a_n x^n$ and
$g(x) = b_s x^s + \cdots + b_t x^t$
be two elements of $F[x, x^{-1}]$ with a_m, a_n, b_s, b_t all non-zero. Then $N(f) = n - m$, and since $a_n b_t \neq 0$ and $a_m b_s \neq 0$, we have that

$$N(fg) = (n + t) - (m + s) = (n - m) + (t - s) \geq N(f).$$

Thus condition (i) holds.

If $N(g) > N(f)$ then condition (ii) holds with $q = 0$ and $r = f$. We now show that condition (ii) also holds in the case $N(f) \geq N(g)$, that is $n - m \geq t - s$.

Consider $x^{-m} f = a_m + a_{m+1} x + \cdots + a_n x^{n-m}$ and
$x^{-s} g = b_s + b_{s+1} x + \cdots + b_t x^{t-s}$.
These are both polynomials in $F[x]$, which is a Euclidean domain. Therefore there are polynomials $q_1, r_1 \in F[x]$ such that

$$x^{-m} f = (x^{-s} g) q_1 + r_1,$$

where $r_1 = 0$ or $\deg(r_1) < \deg(x^{-s} g)$.

Then $f = x^{m-s} g q_1 + x^m r_1$. Let $q = x^{m-s} q_1$ and $r = x^m r_1$. Then:

- if $r_1 = 0$, then $r = 0$ and condition (ii) holds;
- otherwise, we know $\deg(r_1) < \deg(x^{-s} g)$, so $N(r_1) < N(x^{-s} g)$ in $F[x, x^{-1}]$. But $N(r_1) = N(r)$ and $N(x^{-s} g) = N(g)$ by what you observed after the definition of N. Therefore $N(r) < N(g)$, hence condition (ii) holds.

Since N is a Euclidean norm for $F[x, x^{-1}]$, $F[x, x^{-1}]$ is a Euclidean domain and hence a UFD.

2 Ideals

This section discusses ideals, which are special subsets of rings that play a similar role in ring theory as normal subgroups do in group theory. For example, the quotient of a ring by one of its ideals is again a ring, with suitably defined operations between the cosets of an ideal.

2.1 Ideals and their properties

In Book A, Chapter 3, Subsection 1.2 we described how the modular rings $\mathbb{Z}_n$ are constructed out of the ring $\mathbb{Z}$ by splitting $\mathbb{Z}$ up into the set of disjoint residue classes $\{[0], [1], \cdots, [n-1]\}$. For example, the elements of $\mathbb{Z}_3$ are the residue classes

$$[0] = \{\ldots, -6, -3, 0, 3, 6, \ldots\},$$
$$[1] = \{\ldots, -5, -2, 1, 4, 7, \ldots\},$$
$$[2] = \{\ldots, -4, -1, 2, 5, 8, \ldots\}.$$

Normally $\mathbb{Z}_3$ is represented by a complete set of residues modulo 3 such as $\{0, 1, 2\}$. We also showed that the operations of addition, subtraction and multiplication on residue classes are well defined by the rules

$$[a] \pm [b] = [a \pm b] \qquad \text{and} \qquad [a] \times [b] = [a \times b].$$

In this section we begin to explore whether this process can be generalised to rings other than $\mathbb{Z}$. Can we partition any ring R into subsets that form the elements of a new ring? If so, what does the zero element of such a ring look like? Two observations about the zero element $[0]$ of $\mathbb{Z}_n$ can guide us here. The first is that if we subtract $[0]$ from itself in $\mathbb{Z}_n$ then we must again obtain $[0]$. The second observation is that if we multiply $[0]$ by any element $[c] \in \mathbb{Z}_n$, we must also obtain $[0]$. These seemingly trivial observations tell us that the set $[0]$ must be closed under subtraction, and also under multiplication by elements of $\mathbb{Z}$. For if $c \in \mathbb{Z}$ and $a, b \in [0]$ then $[a] = [b] = [0]$, and so our two observations yield

$$[a - b] = [a] - [b] = [0] \qquad \text{and} \qquad [a \times c] = [a] \cdot [c] = [0].$$

Hence $a - b \in [0]$ and $a \cdot c \in [0]$, as suggested.

Subsets of a ring that have these two closure properties play a pivotal role in generalising the modular construction to arbitrary rings. They are called *ideals*.

Definition 2.1 *Ideal*

Let $(R, +, \cdot)$ be a commutative ring, and let I be a non-empty subset of R. Then I is an **ideal** of R if

I1 for all $a, b \in I$, $a - b \in I$

I2 for all $a \in I$ and $r \in R$, $ra \in I$.

Every ring R has at least two ideals, namely $\{0\}$ and R. It is sometimes convenient to be able to emphasise whether or not an ideal is one of these.

Definition 2.2 *Proper and non-trivial ideal*

An ideal I of a ring R is said to be **proper** if it is not equal to R. It is said to be **non-trivial** if it is not equal to $\{0\}$.

Any ideal I that contains the multiplicative identity 1 cannot be a proper ideal. This is because if $1 \in I$ and $a \in R$, then $a = a \cdot 1 \in I$ by I2. So $I = R$. It follows that a proper ideal of R cannot be a subring of R because it does not contain 1. Conversely, any ideal that does not contain 1 is proper.

By contrast, every ideal I contains the additive identity 0. Indeed, since I is non-empty it contains an element a, so $0 = a - a \in I$ by I1.

We emphasise these general properties of ideals below.

Let I be an ideal of the ring R. Then:

- $0 \in I$
- I is a proper ideal if, and only if, it does not contain the multiplicative identity 1.

Exercise 2.1

Let I be an ideal of a ring R and suppose that $a, b \in I$. Show that $a + b \in I$.

If we ignore the operation of multiplication on R for a moment, then we are left with a commutative group $(R, +)$, of which I is a normal subgroup by I1. In Section 3, we will show how the multiplication operation on R can be used to define a multiplication operation on the quotient group R/I, thereby turning it into a *quotient ring*. This will be the generalisation of the modular ring $\mathbb{Z}_n$ that we have been seeking. For now, however, we concentrate on exploring some further examples and properties of ideals.

Worked Exercise 2.3

Show that the set $I = \{f \in \mathbb{Z}[x] : f(3) = 0\}$ is an ideal of $\mathbb{Z}[x]$.

Solution

Since I contains the zero polynomial, it must be non-empty.

Moreover,

I1 if $f, g \in I$, then $(f-g)(3) = f(3) - g(3) = 0$, and so $f - g \in I$;

I2 if $f \in I$ and $r \in \mathbb{Z}[x]$, then $(rf)(3) = r(3)f(3) = r(3) \cdot 0 = 0$, so $rf \in I$.

So the ideal conditions are satisfied, and I is an ideal of $\mathbb{Z}[x]$.

Exercise 2.2

Let R be a commutative ring, and let $r \in R$. Show that the set

$$K = \{x \in R : xr = 0\}$$

is an ideal of R.

In the next exercise, we verify that the zero residue classes of a modular ring of integers, for example $[0] = \{\ldots, -6, -3, 0, 3, 6, \ldots\}$ in $\mathbb{Z}_3$, are ideals of $\mathbb{Z}$.

Worked Exercise 2.4

Let $n \in \mathbb{Z}$. Show that the set $n\mathbb{Z} = \{nk : k \in \mathbb{Z}\}$ is an ideal of $\mathbb{Z}$.

Solution

We first observe that $0 = n0 \in n\mathbb{Z}$, so $n\mathbb{Z}$ is non-empty.
Next we check the ideal conditions.

I1 Let nk_1 and nk_2 be two arbitrary elements of $n\mathbb{Z}$. Then

$$nk_1 - nk_2 = n(k_1 - k_2) \in n\mathbb{Z}.$$

I2 Let nk be an arbitrary element of $n\mathbb{Z}$ and let $m \in \mathbb{Z}$. Then

$$m(nk) = n(mk) \in n\mathbb{Z}.$$

We are using commutativity in $\mathbb{Z}$ here.

So $n\mathbb{Z}$ is an ideal of $\mathbb{Z}$.

Similarly, in the polynomial ring $\mathbb{Q}[x]$ the sets of all possible multiples of a given polynomial are ideals.

Exercise 2.3

Let q be a polynomial in $\mathbb{Q}[x]$. Show that the set

$$q\mathbb{Q}[x] = \{qk : k \in \mathbb{Q}[x]\}$$

is an ideal of $\mathbb{Q}[x]$.

In general, any set aR obtained by multiplying each element of a commutative ring R by a single element $a \in R$ is an ideal. Such a set is certainly non-empty, because $0 \in aR$, and moreover

I1 if $ar_1, ar_2 \in aR$, then $ar_1 - ar_2 = a(r_1 - r_2) \in aR$;

I2 if $k \in R$ and $ar \in aR$, then $k(ar) = a(kr) \in aR$.

So aR is indeed an ideal.

Such ideals occur sufficiently often to merit a special name.

Definition 2.5 *Principal ideal*

Let R be a commutative ring, and let $a \in R$. Then the set

$$aR = \{ar : r \in R\}$$

is the **principal ideal** of R generated by a. If the ring R is clear from the context, we often write $\langle a \rangle$ instead of aR.

Although in this book all the rings discussed are commutative, we have emphasised that R is commutative here as this is a definition.

Example 2.6 *Some principal ideals*

(a) In any ring R, the ideals generated respectively by 0 and 1 are

$$\langle 0 \rangle = 0\,R = \{0\} \quad \text{and} \quad \langle 1 \rangle = 1\,R = R.$$

(b) The ideal of $\mathbb{Z}$ generated by 5 is

$$\langle 5 \rangle = 5\mathbb{Z} = \{\ldots, -15, -10, -5, 0, 5, 10, 15, \ldots\}.$$

(c) The ideal of $\mathbb{Z}[i] = \{a + ib : a, b \in \mathbb{Z}\}$ generated by 5 is

$$\langle 5 \rangle = 5\mathbb{Z}[i] = \{5a + 5bi : a, b \in \mathbb{Z}\} = \{a + bi : a, b \in 5\mathbb{Z}\}.$$

(d) The ideal of $\mathbb{Z}[x]$ generated by x^2 is

$$\langle x^2 \rangle = x^2\mathbb{Z}[x] = \{a_0x^2 + a_1x^3 + \cdots + a_nx^{n+2} : a_0, a_1, \cdots, a_n \in \mathbb{Z}\}.$$

Exercise 2.4

Let R be a ring and let u be a unit in R. Show that $\langle u \rangle = R$.

The next exercise asks you to show that fields can be characterised among general rings in terms of their ideals.

Exercise 2.5

Show that a commutative ring F is a field if, and only if, it contains no proper non-trivial ideals.

In the next exercise, we use the notation $R[\alpha]$, where R is a subring of a ring S and α is an element of S, to denote the smallest subring of S that

contains R and α. For example, if α is a square root of an element in R, the set

$$\{a + b\alpha : a, b \in R\}$$

is a subring of S that contains both R and α, and you will see that

$$R[\alpha] = \{a + b\alpha : a, b \in R\}.$$

Exercise 2.6

Let R be a subring of S and suppose that $\alpha \in S$ with $\alpha^2 = r \in R$.

(a) Show that $T = \{a + b\alpha : a, b \in R\}$ is a subring of S.

(b) Show that $T = R[\alpha]$.

Exercise 2.7

Let $R = \mathbb{Z}[\sqrt{2}] = \{a + b\sqrt{2} : a, b \in \mathbb{Z}\}$.

(a) Show that $\langle 1 + \sqrt{2} \rangle = R$.

(b) Write down a non-zero element x of R such that $\langle x \rangle \neq R$.

Now consider the following three principal ideals of $\mathbb{Z}$:

$$2\mathbb{Z} = \{\ldots, -8, -6, -4, -2, 0, 2, 4, 6, 8, \ldots\},$$
$$6\mathbb{Z} = \{\ldots, -24, -18, -12, -6, 0, 6, 12, 18, 24, \ldots\},$$
$$12\mathbb{Z} = \{\ldots, -48, -36, -24, -12, 0, 12, 24, 36, 48, \ldots\},$$

and notice that they satisfy $12\mathbb{Z} \subseteq 6\mathbb{Z} \subseteq 2\mathbb{Z}$. Notice, too, that $2 \mid 6$ and $6 \mid 12$. In the next exercise you are asked to show that this is an instance of a general property of the principal ideals of a ring.

Exercise 2.8

Let R be a ring, and let $a, b \in R$. Show that $aR \subseteq bR$ if, and only if, $b \mid a$.

The condition $b \mid a$ is equivalent to the condition $a \in bR$. Both are ways of saying that a is a multiple of b. We summarise these equivalences in the next lemma.

Lemma 2.7

Let R be a commutative ring, and let $a, b \in R$. Then the following statements are equivalent.

(a) $aR \subseteq bR$

(b) $b \mid a$

(c) $a \in bR$

Conditions (a) and (c) can also be written as $\langle a \rangle \subseteq \langle b \rangle$ and $a \in \langle b \rangle$ respectively.

Proof You showed that (a) ⇔(b) in the solution to Exercise 2.8. We are left to show that (b) ⇔ (c).

Suppose that $b \mid a$. Then there is an $s \in R$ such that $a = bs$, so $a \in bR$. Conversely, suppose that $a \in bR$. Then there is an element $s \in R$ such that $a = bs$ and so $b \mid a$. Thus we have shown (b) ⇔ (c) as required. ■

Exercise 2.9

We defined associates in Book C, Chapter 12, Definition 2.3.

Let R be an integral domain, and let $a, b \in R$. Show that $aR = bR$ if, and only if, a and b are associates.

A sequence of ideals such as

$$60\mathbb{Z} \subseteq 12\mathbb{Z} \subseteq 6\mathbb{Z} \subseteq 2\mathbb{Z},$$

in which each ideal is a subset of its successor, is known as a **chain of ideals**. A chain can consist of an infinite sequence of ideals,

$$I_0 \subseteq I_1 \subseteq I_2 \subseteq \cdots .$$

Infinite chains have a simple property that you are asked to explore in the next exercise.

Exercise 2.10

Let R be a ring, and let $I_0 \subseteq I_1 \subseteq I_2 \subseteq, \cdots$ be a chain of ideals of R.

Show that the union I of all the ideals I_n for $n = 0, 1, 2, \ldots$ is itself an ideal of R.

Do not be deceived by this exercise. There is no guarantee that an arbitrary union of ideals is an ideal. The solution to Exercise 2.10 depends crucially on the fact that the ideals form a chain. The next exercise shows this, and it also shows that intersections are better behaved.

Exercise 2.11

(a) Show that if I_1 and I_2 are ideals of a ring R then their intersection $I_1 \cap I_2$ is an ideal of R.

(b) By considering the ideals $I_1 = 2\mathbb{Z}$ and $I_2 = 7\mathbb{Z}$ of $\mathbb{Z}$, show that their union $I_1 \cup I_2$ is not an ideal of R.

Hint: for part (b), consider $2 - 7 = -5$ in $\mathbb{Z}$.

2.2 Principal ideal domains

Up to now, all of the ideals in $\mathbb{Z}$ that we have considered have been principal ideals consisting of all the multiples of some integer n. It seems natural to ask whether there are any ideals in $\mathbb{Z}$ that are not principal. The next theorem assures us that the answer is no.

Theorem 2.8

Every ideal in $\mathbb{Z}$ is a principal ideal. Every non-zero ideal in $\mathbb{Z}$ is generated by the smallest positive integer that it contains.

Proof Let I be an ideal in $\mathbb{Z}$. If $I = \langle 0 \rangle$, I is principal, so we can assume $I \neq \langle 0 \rangle$.

If $I \neq \langle 0 \rangle$, I must contain a positive integer, because it contains an integer m, and $m \in I \implies -m = (-1)m \in I$. Let n be the smallest positive integer in I.

First we show that n divides every element of I, so that $I \subseteq n\mathbb{Z}$. The proof uses the Euclidean Algorithm.

Let a be an arbitrary element of I and suppose that n does not divide a. Then, by the Euclidean Algorithm, there are integers k and r such that $a = kn + r$, where $0 < r < n$. Now, $a \in I$ and $n \in I$, and so $a - kn = r \in I$. But this contradicts the fact that n is the smallest positive integer in I, so it must be the case that n divides a and therefore $I \subseteq n\mathbb{Z}$.

For the reverse inclusion, since I is an ideal in $\mathbb{Z}$ and $n \in I$, it follows from closure under multiplication by elements of $\mathbb{Z}$ that $n\mathbb{Z} \subseteq I$. So $I = n\mathbb{Z}$, and we have shown that I is a principal ideal whose least positive element is a generator. ■

A very similar proof shows that every ideal I in the ring $\mathbb{Q}[x]$ is also principal, and it is generated by a polynomial of least degree in I. The underlying common feature here is that both $\mathbb{Z}$ and $\mathbb{Q}[x]$ are Euclidean domains, and so we can exploit the Euclidean Algorithm.

The following theorem is a generalisation of Theorem 2.8 to all Euclidean domains.

Theorem 2.9

Every ideal I in a Euclidean domain R is a principal ideal generated by an element with the least norm in I, or by 0 in the case $I = \{0\}$.

Proof Let R be a Euclidean domain with norm

$$d : R \setminus \{0\} \to \mathbb{N} \cup \{0\},$$

and let I be an ideal of R. If $I = \langle 0 \rangle$, I is principal, so we can assume $I \neq \langle 0 \rangle$. By the solution of Exercise 2.4, I is also principal if $1 \in I$, for then $I = \langle 1 \rangle = R$.

So let a be an element in I such that $a \neq 0$ and $0 < d(a) \leq d(b)$ for every $b \in I$. We show that a divides every element of I, and so $I \subseteq \langle a \rangle$ by Lemma 2.7.

Let b be an arbitrary element of I and suppose that a does not divide b. Then, by the Euclidean Algorithm, there are elements s and r in R such that $b = sa + r$, where $0 < d(r) < d(a)$.

Now $b \in I$ and $sa \in I$, and so $b - sa = r \in I$. But this contradicts the fact that a is an element in I with the least norm among all elements in I, and so it must be the case that a divides every element of I, and therefore $I \subseteq \langle a \rangle$.

For the reverse inclusion, since I is an ideal in R and $a \in I$, it follows from closure under multiplication that $\langle a \rangle \subseteq I$. So $I = \langle a \rangle$, and we have shown that I is a principal ideal whose generator is an element with least norm. ■

There is a special name for integral domains whose ideals are all principal.

Definition 2.10 *Principal ideal domain*

A **principal ideal domain** (PID) is an integral domain in which every ideal is principal.

The following corollary of Theorem 2.9 provides us with a rich supply of principal ideal domains.

Corollary 2.11

Every Euclidean domain is a principal ideal domain.

Corollary 2.11 tells us, for example, that the integers $\mathbb{Z}$, the Gaussian integers $\mathbb{Z}[i]$, and the ring of polynomials $F[x]$, where F is a field, are all examples of PIDs, for they are all Euclidean domains.

This raises the question as to whether there are any integral domains that have non-principal ideals. By Corollary 2.11, such a domain cannot be Euclidean, so it may be worth searching for an example in the integral domain $\mathbb{Z}[x]$, which was shown not to be Euclidean in Exercise 3.3(b) of Book C, Chapter 12.

Worked Exercise 2.12 *A non-principal ideal*

Let A be the set of polynomials in $\mathbb{Z}[x]$ that have an even constant term. Show that A is an ideal of $\mathbb{Z}[x]$ that is not principal.

Solution

First we observe that $0 \in A$, so A is non-empty. Next let

$$q(x) = 2a + a_1x + \cdots + a_nx^n \quad \text{and}$$
$$r(x) = 2b + b_1x + \cdots + b_mx^m$$

be two arbitrary elements of A, and

$$u(x) = c_0 + c_1x + \cdots + c_kx^k$$

be an arbitrary element of $\mathbb{Z}[x]$. Then

I1 the constant term of $q - r$ is $2(a - b)$, which is even, so $q - r \in A$;

I2 the constant term of uq is $2(c_0a)$, so $uq \in A$.

Thus A is an ideal of $\mathbb{Z}[x]$.

Now suppose that A is a principal ideal generated by a polynomial q. Since $2 \in A$, it follows that $2 = uq$ for some $u \in \mathbb{Z}[x]$. Then q must be a constant polynomial, otherwise
$\deg(uq) = \deg(u) + \deg(q) > 0 = \deg(2)$. But the polynomial x is an element of A, and so $x = vq$ for some $v \in \mathbb{Z}[x]$. Since q is constant, either $v = x$ and $q = 1$, or $v = -x$ and $q = -1$. Either way, q is an odd constant polynomial, which contradicts $q \in A$. So A is not principal.

It should be noted that the converse of Corollary 2.11 is not true: not all principal ideal domains are Euclidean domains. Nevertheless, some of the results that hold for Euclidean domains do carry over to PIDs. For example, highest common factors exist in PIDs.

For example, it can be shown that $\{\frac{1}{2}\left(a + b\sqrt{19}i\right) : a, b \in \mathbb{Z}\}$ is a subring of $\mathbb{C}$ that is a PID but not an ED. The proof, however, is rather technical, and beyond the scope of this module. It can be found in 'A Principal Ideal Domain That Is Not a Euclidean Domain', by Oscar A. Campoli, *The American Mathematical Monthly*, vol. 95, no. 9 (Nov. 1988), pp. 868–71.

Theorem 2.13 *Highest common factors in PIDs*

Let R be a principal ideal domain, and let $a, b \in R$. Then a and b have a highest common factor $\mathrm{hcf}(a, b)$, that is, a common factor that is a multiple of any other common factor.

Moreover, if $h = \mathrm{hcf}(a, b)$ then there are elements $s, t \in R$ such that $h = sa + tb$.

Proof Let $I = \{ua + vb : u, v \in R\}$. We leave you to check that I is an ideal of R. As R is a PID, there is an element $h \in R$ such that $I = \langle h \rangle$. Since $a \in I$ and $b \in I$, it follows that a and b are multiples of h, so $h \mid a$ and $h \mid b$. In other words, h is a common factor of a and b.

Since $h \in I$, there must be elements $s, t \in R$ such that $h = sa + tb$. So if $d \mid a$ and $d \mid b$ for some $d \in R$, then $d \mid (sa + tb)$, that is, $d \mid h$. So h is the highest common factor. Thus $h = \mathrm{hcf}(a, b)$ exists and $h = sa + tb$. ■

Had Worked Exercise 2.12 asked us to show that $\mathbb{Z}[x]$ is not a principal ideal domain, rather than check that A is not a principal ideal of $\mathbb{Z}[x]$, then we could have saved a little work simply by showing that there are no $s, t \in \mathbb{Z}[x]$ for which $\text{hcf}(2, x) = 1 = 2s(x) + xt(x)$, as we did in the solution to Exercise 3.3 of Book C, Chapter 12. The contrapositive form of Theorem 2.13 then confirms that $\mathbb{Z}[x]$ is not a principal ideal domain.

Worked Exercise 2.14

Find a highest common factor of the elements $3 + i$ and $1 + 3i$ in $\mathbb{Z}[i]$.

Solution

A highest common factor exists because $\mathbb{Z}[i]$ is a PID. In fact $\mathbb{Z}[i]$ is a Euclidean domain, so we can use the Euclidean Algorithm to find $\text{hcf}(3 + i, 1 + 3i)$. We begin by using ordinary complex division to obtain

$$\frac{3+i}{1+3i} = \frac{1}{10}(6 - 8i) = (1 - i) + \frac{1}{5}(-2 + i),$$

Multiply the numerator and the denominator by $1 - 3i$.

where $1 - i$ is the element of $\mathbb{Z}[i]$ closest to $\frac{1}{10}(6 - 8i)$. Then multiplying through by $1 + 3i$ gives the quotient–remainder form

$$3 + i = (1 - i)(1 + 3i) + (-1 - i).$$

An element of $\mathbb{Z}[i]$ is 'closest' to another in the sense of the Euclidean norm $N(a + bi) = a^2 + b^2$.

Since there is still a non-zero remainder, we calculate

$$\frac{1+3i}{-1-i} = -2 - i \in \mathbb{Z}[i].$$

The remainder is now zero, so an hcf is the previous remainder:

$$\text{hcf}(3 + i, 1 + 3i) = -1 - i.$$

Since -1 is a unit, $1 + i$ is also an hcf.

Exercise 2.12

Find a highest common factor, if one exists, of the elements $(9 + 11\sqrt{-2})$ and $(1 + 5\sqrt{-2})$ in $\mathbb{Z}[\sqrt{-2}]$.

We can use Theorem 2.13 to establish another result that carries over from Euclidean domains to PIDs.

Theorem 2.15

Let R be a principal ideal domain. Then p is prime in R if, and only if, p is an irreducible element of R.

Proof Since in any integral domain a prime is irreducible (see Book C, Chapter 12, Theorem 2.12), we are left to show that any irreducible element of R is prime.

Let $p \in R$ be irreducible, and suppose $p \mid ab$. Let $h = \text{hcf}(p, a)$ and write

$$h = up + va$$

for some $u, v \in R$.

If h is an associate of p, then $p \mid h$, and since $h \mid a$ it follows that $p \mid a$. If, on the other hand, h is not an associate of p, then the only way that h can divide p is if $p = rh$ for some non-unit $r \in R$. Since p is irreducible, this implies that h is a unit. We can therefore write

$$1 = h^{-1}(up + va),$$

and hence $b = (h^{-1}ub)p + (h^{-1}v)ab$. So b is divisible by p. In either case, p divides a or b, and so p is prime in R. ■

We can also show that every principal ideal domain is a unique factorisation domain (UFD).

Theorem 2.16 *Unique factorisation in PIDs*

Every principal ideal domain R is a unique factorisation domain.

Proof Let a_0 be any non-trivial element of R, where by 'non-trivial' we mean an element that is neither a unit nor zero. Suppose that a_0 cannot be expressed as a product of irreducible elements. Then a_0 cannot itself be irreducible, so $a_0 = a_1 b_1$ for some non-trivial $a_1, b_1 \in R$, at least one of which (a_1 say) cannot be expressed as a product of irreducibles. Then $a_1 = a_2 b_2$ for some non-trivial $a_2, b_2 \in R$, at least one of which (a_2 say) cannot be expressed as a product of irreducibles. Continuing in this way, we obtain a sequence

$$a_0, a_1, a_2, \ldots$$

where each element is a divisor of its predecessor. So by Lemma 2.7 we can form the chain

$$\langle a_0 \rangle \subseteq \langle a_1 \rangle \subseteq \langle a_2 \rangle \subseteq \cdots$$

of ideals in R. By the solution to Exercise 2.10, the union of all these ideals must itself be an ideal and it must therefore be equal to some principal ideal $\langle a \rangle$. Since a is an element of this union, it must also be an element of $\langle a_i \rangle$ for some i. So $\langle a \rangle \subseteq \langle a_i \rangle \subseteq \langle a_{i+1} \rangle \subseteq \langle a \rangle$ and we conclude that $\langle a_i \rangle = \langle a_{i+1} \rangle$. But this implies that a_i and a_{i+1} are associates by the solution to Exercise 2.9. Since this contradicts our previous statement that b_{i+1} is not a unit, we conclude that a_0 is in fact a product of irreducible elements.

The result now follows by Theorem 2.15 and Theorem 4.2 of Book C, Chapter 12. ■

We know from Theorem 4.5 and Corollary 4.6 of Book C, Chapter 12, that $\mathbb{Z}[x]$ is a UFD. However, you saw in Worked Exercise 2.12 that $\mathbb{Z}[x]$ is not a PID. We may therefore conclude that the converse of Theorem 2.16 is false: not all UFDs are PIDs.

Exercise 2.13

Decide whether or not each of the following rings is a PID, giving reasons for your answer.

(a) $\mathbb{Z}[\sqrt{-2}] = \{a + b\sqrt{-2} : a, b \in \mathbb{Z}\}$

(b) $\mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} : a, b \in \mathbb{Z}\}$

We close this section with a summary of the key results about various types of integral domain, and prime and irreducible elements. Figure 2.1 then shows the relationships between the various types of commutative rings we have discussed so far. It also gives examples that show that the inclusions are proper.

Results about different types of domain

- A polynomial ring in one variable over a field is a Euclidean domain (Book C, Chapter 11, Theorem 3.1).
- Every Euclidean domain is a principal ideal domain (Corollary 2.11).
- Every principal ideal domain is a unique factorisation domain (Theorem 2.16).
- If R is a unique factorisation domain then $R[x]$ is a unique factorisation domain (Theorem 1.16).
- Unique factorisation domains need not be Euclidean domains (Book C, Chapter 12, Example 4.5).
- Not all principal ideal domains are Euclidean domains (example cited in the margin note preceding Theorem 2.13).
- Not all unique factorisation domains are principal ideal domains. For example, $\mathbb{Z}[x]$ is a unique factorisation domain that is not a principal ideal domain.

Results about irreducible elements

- Let R be an integral domain and p a prime element in R. Then p is irreducible (Book C, Chapter 12, Theorem 2.12).
- Let R be a unique factorisation domain. Then p is prime in R if, and only if, p is an irreducible element of R (Book C, Chapter 12, Theorem 2.12 and Corollary 4.3).

- Let R be an integral domain in which every non-zero non-unit element has a factorisation into irreducible elements. Then R is a unique factorisation domain if, and only if, every irreducible element in R is prime (Book C, Chapter 12, Theorem 4.2).
- Let R be a unique factorisation domain and let $a, b \in R$. Then a and b have a highest common factor $\text{hcf}(a, b)$ (Book C, Chapter 12, Proposition 4.7).
- Let R be a principal ideal domain and let $a, b \in R$. Then a and b have a highest common factor $\text{hcf}(a, b)$. If $h = \text{hcf}(a, b)$, then there are elements $s, t \in R$ such that $h = sa + tb$ (Theorem 2.13).

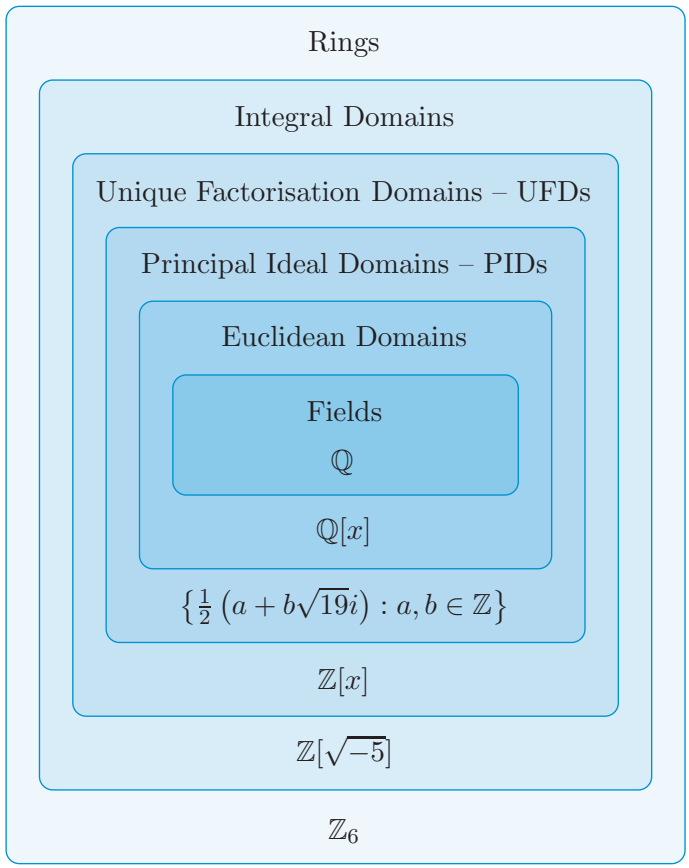

Figure 2.1 The relationships between different types of commutative rings

3 Operations on ideals

In Exercise 2.11(a) you saw that the intersection of two ideals is an ideal. This section explores other natural ways of obtaining new ideals from existing ones.

3.1 Sum of ideals

The next lemma will allow us to define an operation of addition between ideals.

Lemma 3.1

Let A and B be ideals of a ring R. Then the set

$$A + B = \{a + b : a \in A, b \in B\}$$

is an ideal of R.

Proof Since $0 + 0 = 0$, $A + B$ contains 0 and so it is non-empty. It remains to check conditions I1 and I2 of Definition 2.1.

I1 Let $a + b$ and $a_1 + b_1$ be arbitrary elements of $A + B$. Then

$$(a + b) - (a_1 + b_1) = (a - a_1) + (b - b_1) \in A + B.$$

I2 Let $a + b$ be an arbitrary element of $A + B$, and let $r \in R$. Then

$$r(a + b) = (ra) + (rb) \in A + B.$$

So $A + B$ is an ideal. ■

We can now define the sum of two ideals.

Definition 3.2 *Sum of two ideals*

The **sum of two ideals** A and B of a ring R is the ideal

$$A + B = \{a + b : a \in A, b \in B\}.$$

Exercise 3.1

Let A and B be ideals of a commutative ring R. Prove that

$$A \subseteq (A + B) \text{ and } B \subseteq (A + B).$$

Notice that since $A + B$ contains A and B, it must in fact be the smallest ideal containing A and B. To see this, let K be an ideal containing both A and B, and let $a + b \in A + B$ with $a \in A$ and $b \in B$. Then $a + b \in K$ by the solution to Exercise 2.1.

We now list some of the basic properties of addition of ideals.

Proposition 3.3 *Basic additive properties of ideals*

Let A, B and C be ideals of a commutative ring R. Then:

(a) $A \subseteq (A+B)$ and $B \subseteq (A+B)$

(b) $A+B = B+A$

(c) $(A+B)+C = A+(B+C)$

(d) $A+\{0\} = A$

(e) $A+R = R$.

Proof

(a) See the solution to Exercise 3.1.

(b) $A+B = \{a+b : a \in A, b \in B\}$
$= \{b+a : b \in B, a \in A\} = B+A$

(c) $(A+B)+C = \{u+c : u \in (A+B), c \in C\}$
$= \{(a+b)+c : a \in A, b \in B, c \in C\}$
$= \{a+(b+c) : a \in A, b \in B, c \in C\}$
$= \{a+v : a \in A, v \in (B+C)\}$
$= A+(B+C)$

(d) $A+\{0\} = \{a+0 : a \in A\} = A$

(e) By part (a), $R \subseteq A+R$, so $A+R = R$. ■

Exercise 3.2

In a principal ideal domain R, show that $\langle a \rangle + \langle b \rangle = \langle h \rangle$, where $h = \mathrm{hcf}(a,b)$.

3.2 Product of ideals

Next we consider how to define the product of two ideals I and J. By analogy with the definition of sum, the obvious candidate might appear to be the set of products $\{ab : a \in I, b \in J\}$. The snag is that this set is not usually closed under addition, and so it is not necessarily an ideal. The answer is to use arbitrary finite sums of products of an element of I and an element of J, that is, elements of the form

$$a_1b_1 + a_2b_2 + \cdots + a_nb_n,$$

in which any finite number n of terms is allowed.

Exercise 3.3

Let A and B be ideals in a ring R. Show that the set

$$AB = \{a_1b_1 + a_2b_2 + \cdots + a_nb_n : n \in \mathbb{N} \text{ and } a_i \in A, b_i \in B\},$$

is an ideal of R.

The solution to this exercise ensures that we can define the product of two ideals.

Definition 3.4 *Product of two ideals*

The **product of two ideals** A and B of a ring R is the ideal

$$AB = \{a_1b_1 + a_2b_2 + \cdots + a_nb_n : n \in \mathbb{N} \text{ and } a_i \in A, b_i \in B\}.$$

The sum and product share some of the same properties. They are, for example, both commutative and associative. However, there are also some differences. While the sum $A + B$ contains A and B, the product AB is contained in A and B. Also, the way in which the ideals $\{0\}$ and R combine with other ideals is very different.

Proposition 3.5 *Basic multiplicative properties of ideals*

Let A, B and C be ideals of a commutative ring R. Then:

(a) $AB \subseteq A$ and $AB \subseteq B$

(b) $AB = BA$

(c) $(AB)C = A(BC)$

(d) $A\{0\} = \{0\}$

(e) $AR = A$.

Proof In what follows we, adopt the convention that a, b, c (with or without subscripts) represent arbitrary elements of A, B, C, respectively.

(a) Let $x \in AB$. Then x is a sum of terms of the form ab, where $a \in A$ and $b \in B$. Each such ab belongs to both A and B, because ideals are closed under multiplication by elements of R, so x belongs to A and B.

That is, $AB \subseteq A$ and $AB \subseteq B$.

(b) Each element of AB has the form

$$a_1b_1 + \cdots + a_kb_k = b_1a_1 + \cdots + b_ka_k \in BA,$$

and each element of BA has the form

$$b_1a_1 + \cdots + b_ka_k = a_1b_1 + \cdots + a_kb_k \in AB.$$

So $AB = BA$.

(c) Let $x \in (AB)C$. Then x is a sum of elements of the form

$$(a_1b_1 + \cdots + a_kb_k)c = a_1(b_1c) + \cdots + a_k(b_kc),$$

so $x \in A(BC)$.

Now let $x \in A(BC)$. Then x is a sum of elements of the form

$$a(b_1c_1 + \cdots + b_\ell c_\ell) = (ab_1)c_1 + \cdots + (ab_\ell)c_\ell,$$

so $x \in (AB)C$. Hence $(AB)C = A(BC)$.

(d) By part (a), $A\{0\} \subseteq \{0\}$, so $A\{0\} = \{0\}$.

(e) Let $x \in A$. Then $x = x \cdot 1 \in AR$. So $A \subseteq AR$. But we also have $AR \subseteq A$ by part (a), hence $AR = A$. ■

3.3 Finitely generated ideals

Just as we defined the principal ideal generated by a single element, we now define the ideal generated by a finite subset $S = \{a_1, \ldots, a_n\}$ of a ring R. You met an ideal generated by two elements in the proof of Theorem 2.13, where you were left to check that

$$\{ua + vb : u, v \in R\}$$

is an ideal for two given elements a and b of a ring R. The next definition generalises this way to define an ideal.

Definition 3.6 *Ideal generated by a set*

Let R be a commutative ring. Then the **ideal of R generated by a finite subset** $S = \{a_1, \ldots, a_n\}$ of R is the set $\langle a_1, \ldots, a_n \rangle$ defined by

$$\langle a_1, \ldots, a_n \rangle = \{a_1r_1 + \cdots + a_nr_n : r_1, \ldots, r_n \in R\}.$$

The notation $\langle a_1, \ldots, a_n \rangle$ is sometimes abbreviated to $\langle S \rangle$.

The set $\langle a_1, \ldots, a_n \rangle$ is an ideal of R because it is the sum of n principal ideals:

$$\langle a_1, \ldots, a_n \rangle = a_1R + a_2R + \cdots + a_nR.$$

In fact, this sum is sometimes used to denote $\langle a_1, \ldots, a_n \rangle$ in cases where the ring under discussion needs to be highlighted.

Note that any given ideal may have many generating sets. In particular, although the ideal $\langle a_1, \ldots, a_n \rangle$ has more than one generator, it may or may not be principal. This fact is illustrated in the next exercise.

Exercise 3.4

Decide whether each of the following ideals is principal. If it is, find an element of the ring that generates it.

(a) The ideal $\langle 27, 36\rangle$ of the ring $\mathbb{Z}$.

(b) The ideal $\langle x, 2\rangle$ of the ring $\mathbb{Z}[x]$.

The approach used to find the generator of the ideal $\langle 27, 36\rangle$ in this exercise can be generalised to any PID.

Lemma 3.7

Let R be a PID, and let $a_1, a_2, \ldots, a_n \in R$. Then

$$\langle a_1, a_2, \ldots, a_n\rangle = \langle \mathrm{hcf}(a_1, a_2, \ldots, a_n)\rangle.$$

Proof The ideal $A = \langle a_1, a_2, \ldots, a_n\rangle$ is principal because R is a PID.

So let h be a generator of A. Then h is a common factor of every element of A, and in particular of the generators $a_1, a_2, \ldots, a_n$.

But $h \in A$, so $h = r_1a_1 + \cdots + r_na_n$ for some $r_1, r_2, \ldots, r_n \in R$. Therefore every common factor of $a_1, a_2, \ldots, a_n$ is also a factor of h.

In other words, $h = \mathrm{hcf}(a_1, a_2, \ldots, a_n)$. ■

Another approach to determining a generator for a principal ideal (or showing that one does not exist) is to use a multiplicative norm function to narrow down, or exclude, possible generators. For example, in $\mathbb{Z}[\sqrt{d}]$ we could take $N(a + b\sqrt{d}) = a^2 - db^2$ as such a norm. Then if an element of the ideal has norm 10, there is no point in looking for a generator that has norm 3, because $3 \nmid 10$.

Exercise 3.5

Decide whether or not each of the following ideals is principal. If it is, find an element of the ring that generates it.

(a) The ideal $A = \langle 3 + i, 1 + 3i\rangle$ of the ring $\mathbb{Z}[i]$.

(b) The ideal $A = \langle 2, 1 + \sqrt{-5}\rangle$ of the ring $\mathbb{Z}[\sqrt{-5}]$.

Hint: for part (b), use a multiplicative norm on $\mathbb{Z}[\sqrt{-5}]$.

The arguments in the solution to Exercise 3.5 are worth generalising. We do so in the next proposition.

Proposition 3.8

Let R be a commutative ring and suppose that $a_1, a_2, \ldots, a_n \in R$. Let A be an ideal of R. Then

$$\langle a_1, a_2, \ldots, a_n \rangle \subseteq A \text{ if, and only if, } a_1, a_2, \ldots, a_n \in A.$$

Proof Certainly $\langle a_1, a_2, \ldots, a_n \rangle \subseteq A$ implies that $a_1, a_2, \ldots, a_n \in A$.

Conversely, $a_1, a_2, \ldots, a_n \in A$ implies that $a_1R + a_2R + \cdots + a_nR \in A$, and so $\langle a_1, a_2, \ldots, a_n \rangle \subseteq A$. ■

Lemma 3.7 and Proposition 3.8 enable us to compare ideals by focusing on their generators. For example, Proposition 3.8 tells us that if we want to show that one ideal A is contained in another ideal B, then it is sufficient to show that the generators of A belong to B.

Exercise 3.6

(a) In the ring $\mathbb{Z}[x]$, show that $\langle x^2 - 1, x^2 - 2x + 1 \rangle$ is a subset of $\langle x - 1 \rangle$.

(b) In the ring $\mathbb{Z}[\sqrt{-5}]$, show that

(i) $\langle 2, 1 + \sqrt{-5} \rangle = \langle 2, 1 - \sqrt{-5} \rangle$

(ii) $\langle 9 \rangle \subseteq \langle 2 - \sqrt{-5} \rangle$.

The following result is useful, and it can be generalised to ideals with any finite number of generators.

Lemma 3.9

The product of the ideals $A = \langle a_1, a_2 \rangle$ and $B = \langle b_1, b_2 \rangle$ of a ring R is the ideal

$$I = \langle a_1b_1, a_1b_2, a_2b_1, a_2b_2 \rangle.$$

Proof The inclusion $I \subseteq AB$ follows from the definition of the product of two ideals and from Proposition 3.8.

The inclusion $AB \subseteq I$ follows from the fact that every element of the product AB is a sum of multiples of the elements $a_1b_1, a_1b_2, a_2b_1, a_2b_2$. ■

In particular, the product of an ideal of the form $\langle a, b \rangle$ with itself is the ideal $\langle a^2, ab, b^2 \rangle$. This is sometimes called the **square** of the ideal.

Worked Exercise 3.10

Let I be the ideal $\langle 2, 1+\sqrt{-5}\rangle$ in the ring $R = \mathbb{Z}[\sqrt{-5}]$. Show that

$$I^2 = 2R.$$

Solution

Since $(1+\sqrt{-5})^2 = -4+2\sqrt{-5}$, we have

$$I^2 = \langle 4, 2+2\sqrt{-5}, -4+2\sqrt{-5}\rangle.$$

Each of these three generators of I^2 can be shown to be in $2R$, hence $I^2 \subseteq 2R$ by Proposition 3.8.

Conversely, since $2R$ is generated by 2 and

$$2 = -4 + (2+2\sqrt{-5}) - (-4+2\sqrt{-5}),$$

we have that $2 \in I^2$, and so $2R \subseteq I^2$.

Therefore $I^2 = 2R$, as required.

Exercise 3.7

Let I be an ideal in a ring R and let $t \in R \setminus I$. Show that the set

$$\langle I, t\rangle = \{m + bt : b \in R, m \in I\}$$

is an ideal of R.

4 Quotients

Our initial reason for introducing ideals was to generalise the way in which the ring $\mathbb{Z}_n$ is constructed by partitioning the ring of integers $\mathbb{Z}$ into disjoint residue classes. We are now ready for this generalisation.

4.1 Cosets of an ideal

Any ring $(R, +, \cdot)$ can be regarded as an abelian group $(R, +)$ together with some extra axioms that govern the behaviour of the product. Each ideal I in R is a subgroup of $(R, +)$, because it is a non-empty set that is closed under subtraction (see Book B, Chapter 5, Proposition 2.5). Recall that any subgroup of an abelian group is normal. Therefore, an ideal of R is a normal subgroup of $(R, +)$. In Book B, Chapter 5, Subsection 2.2 you learnt that the cosets of a subgroup partition the host group into disjoint subsets. This is true in particular for any subgroup of $(R, +)$, but here we are interested only in those subgroups that are ideals, so we give the following definition.

Definition 4.1 *Cosets of an ideal*

Let I be an ideal of a ring R, and let $a \in R$. Then

$$a + I = \{a + r : r \in I\}$$

is the **coset** of I by a.

Note that we have used the coset notation for abelian groups here. This is appropriate since I is a subgroup of the abelian group $(R, +)$. This immediately gives us some useful results from group theory.

Proposition 4.2

Let R be a ring and I an ideal of R. Then the cosets of I partition R into disjoint subsets, and the following four statements are equivalent.

(a) $b - a \in I$

(b) $a + I = b + I$

(c) a and b are in the same coset of I

(d) $b \in a + I$

Proof The required results follow from Propositions 2.12 and 2.13 of Book B, Chapter 5. ■

We illustrate the equivalences in Proposition 4.2 in the following familiar example.

Example 4.3 *Cosets of an ideal in $\mathbb{Z}$*

The set $5\mathbb{Z}$ is an ideal in $\mathbb{Z}$, and its cosets partition $\mathbb{Z}$ into five disjoint subsets as follows.

$$\begin{aligned}
5\mathbb{Z} &= \{\ldots, -10, -5, 0, 5, 10, \ldots\} \\
1 + 5\mathbb{Z} &= \{\ldots, -9, -4, 1, 6, 11, \ldots\} \\
2 + 5\mathbb{Z} &= \{\ldots, -8, -3, 2, 7, 12, \ldots\} \\
3 + 5\mathbb{Z} &= \{\ldots, -7, -2, 3, 8, 13, \ldots\} \\
4 + 5\mathbb{Z} &= \{\ldots, -6, -1, 4, 9, 14, \ldots\}
\end{aligned}$$

We can see from this list that 2 and 3 are in different cosets, and that

$$\begin{aligned}
3 - 2 = 1 &\notin 5\mathbb{Z}, \\
2 + 5\mathbb{Z} &\neq 3 + 5\mathbb{Z}, \\
2 &\notin 3 + 5\mathbb{Z}.
\end{aligned}$$

On the other hand, 3 and -7 are in the same coset, and

$$\begin{aligned}
3 - (-7) = 10 &\in 5\mathbb{Z}, \\
3 + 5\mathbb{Z} &= -7 + 5\mathbb{Z}, \\
-7 &\in 3 + 5\mathbb{Z}.
\end{aligned}$$

Our discussion so far has depended solely on the additive structure of the ring R. However, our aim is not to construct a quotient group from the set of the cosets as we did in the group theory chapters, but rather to construct a quotient *ring*.

While a quotient group requires a single operation to be defined on the set of cosets, a quotient ring requires two, namely addition and multiplication. We already know how to define addition from group theory: if $a+I$ and $b+I$ are two arbitrary cosets of I by a and b respectively, then

$$(a+I)+(b+I)=(a+b)+I.$$

The definition of multiplication is as expected:

$$(a+I)(b+I)=(ab)+I.$$

Before continuing, we must check that this operation of multiplication is well defined. So suppose that $a+I=c+I$ and $b+I=d+I$. Then by Proposition 4.2 we have $a-c\in I$ and $b-d\in I$, so

$$ab-cd=(a-c)b+c(b-d)\in I,$$

using the fact that I is closed under multiplication by any ring elements (here c and b). It follows that $ab+I=cd+I$, and so multiplication is well defined.

We are now close to our goal of defining a quotient ring. You are asked to do some preliminary work in the next exercise.

Exercise 4.1

Let R be a ring, and I be an ideal of R. Show that the set of cosets

$$S=\{a+I : a\in R\}$$

forms a ring under the operations of coset addition $+$ and coset multiplication $\cdot$ defined above.

Our conclusions about the ring $(S,+,\cdot)$ are summarised in the next theorem.

Theorem 4.4

Let I be an ideal of a ring R. Then the set of cosets of I in R, with the operations $+$ and $\cdot$ defined by

$$(a+I)+(b+I)=(a+b)+I \quad \text{and}$$
$$(a+I)\cdot(b+I)=ab+I$$

is a ring, called the **quotient ring** of R by I, and denoted by R/I.

The additive identity of R/I is $0+I$, the multiplicative identity is $1+I$, and the additive inverse of $a+I$ is $-a+I$.

In deriving the above result we have made extensive use of the fact that an ideal of a ring R is also a subgroup of the additive group $(R, +)$. It is very important to realise that the converse is not true: an additive subgroup of R is not necessarily an ideal of R, as the following example illustrates.

Example 4.5 *A subgroup that is not an ideal*

Let $R = \mathbb{Z}[i]$ be the ring of Gaussian integers. Then $\mathbb{Z}$ is a subgroup of $(R, +)$, but $\mathbb{Z}$ is not an ideal of R. This is because any ideal of R that contains the identity element 1 is equal to R, so $\mathbb{Z}$ cannot be an ideal of $\mathbb{Z}[i]$.

We illustrate the process of forming the quotient of a ring through the example of a quotient of the polynomial ring $\mathbb{Q}[x]$.

Worked Exercise 4.6

Let $R = \mathbb{Q}[x]$, and let I be the ideal $\langle x^2 + 1 \rangle$ of R.

(a) Describe the set of cosets R/I.

(b) Find formulas for the coset operations of addition and multiplication.

Solution

(a) Each coset of I has the form $f(x) + I$, where $f \in \mathbb{Q}[x]$. By the Division Algorithm we can write

$$f(x) = q(x)(x^2 + 1) + r(x),$$

where $q, r \in \mathbb{Q}[x]$ and $\deg(r) < \deg(x^2 + 1) = 2$. So

$$f(x) - r(x) = q(x)(x^2 + 1) \in I.$$

Then, by Proposition 4.2,

$$f(x) + I = r(x) + I.$$

But $\deg(r) < 2$, so $r(x)$ has the form $a + bx$, and

$$R/I = \{(a + bx) + I : a, b \in \mathbb{Q}\}.$$

Moreover, $(a + bx) + I = (c + dx) + I$ gives

$$(a + bx) - (c + dx) = (a - c) + (b - d)x \in I,$$

and this is only possible if $a = c$ and $b = d$.

(b) If $a + bx + I$ and $c + dx + I$ are arbitrary elements of R/I, then

$$\begin{aligned}(a + bx + I) + (c + dx + I) &= (a + bx) + (c + dx) + I \\ &= (a + c) + (b + d)x + I.\end{aligned}$$

The product introduces a term in x^2 that we can absorb into I:

$$\begin{aligned}(a+bx+I)\cdot(c+dx+I) &= (a+bx)(c+dx)+I\\ &= ac+(ad+bc)x+bdx^2+I\\ &= (ac-bd)+(ad+bc)x\\ &\quad +bd(x^2+1)+I\\ &= (ac-bd)+(ad+bc)x+I.\end{aligned}$$

Alternatively, we could have gone directly from the second line to the fourth by replacing x^2 by -1. This works because $x^2+1\in I$, and so x^2+1, or rather x^2+1+I, is the additive identity in the quotient ring R/I.

It may have struck you that the formulas for multiplying and adding the cosets in Worked Exercise 4.6 resemble the formulas for adding and multiplying complex numbers. Indeed, if we ignore every '$+I$' and replace every x by i, we end up with valid formulas for adding and multiplying the elements of $\mathbb{Q}[i]$, namely

$$(a+bi)+(c+di)=(a+c)+(b+d)i \quad \text{and}$$
$$(a+bi)\cdot(c+di)=(ac-bd)+(ad+bc)i.$$

In this sense, we can say that $\mathbb{Q}[x]/I$ and $\mathbb{Q}[i]$ look like the same ring. There is a rigorous way to express this resemblance mathematically, and we will discuss this in Subsection 4.3. In the meantime, here is a similar quotient for you to work out.

Exercise 4.2

Let $R=\mathbb{Q}[x]$, and let I be the ideal $\langle x^2+x+1\rangle$ of R.

(a) Describe the set of cosets R/I.

(b) Find formulas for the coset operations of addition and multiplication.

In the solution to Exercise 4.2, the underlying set of the quotient ring $\mathbb{Q}[x]/I$ is $\{(a+bx)+I : a,b\in\mathbb{Q}\}$. This underlying set resembles that of many other rings in this module, for example $\mathbb{Q}[i], \mathbb{Q}[\sqrt{-2}]$ and $\mathbb{Q}[\sqrt{-5}]$. However, none of these rings has a similar multiplicative operation to the one we have found, so any further description of a similarity with a quotient of $\mathbb{Q}[x]$ is not straightforward. We will look at quotients of this kind in much more detail in Chapter 18.

4.2 Homomorphisms

In the paragraph following Worked Exercise 4.6, we remarked that the quotient ring $\mathbb{Q}[x]/\langle x^2+1\rangle$ looks like the ring $\mathbb{Q}[i]$. By contrast, the solution to Exercise 4.2 suggested that the quotient ring $\mathbb{Q}[x]/\langle x^2+x+1\rangle$ is harder to identify. To help us make more systematic comparisons between rings, we introduce ring homomorphisms, which are functions that preserve the ring operations.

Definition 4.7 *Ring homomorphism*

Let R and S be rings. A **ring homomorphism** is a function $f\colon R \to S$ such that, for all $a, b \in R$:

$$f(a+b) = f(a) + f(b)$$
$$f(ab) = f(a)f(b)$$
$$f(1_R) = 1_S,$$

where 1_R and 1_S are the multiplicative identities in R and S respectively.

When S is an integral domain, the condition $f(1_R) = 1_S$ follows from the first two, and may be omitted.

When the fact that f is a function between rings is clear from the context, we often say that f is a homomorphism, rather than a ring homomorphism. We often refer to the two properties in Definition 4.7 as the *homomorphism* or *morphism properties* for addition and multiplication, respectively. This is consistent with the terminology introduced after Definition 1.4.

A ring homomorphism may or may not be one–one, and it may or may not be onto. The following examples illustrate some of the possibilities.

Example 4.8 *Some ring homomorphisms*

(a) Let R be a ring, and let $f\colon R \to R$ be the identity function $f(x) = x$. Then f is a ring homomorphism that is both one–one and onto.

(b) The function $f\colon \mathbb{Z} \to \mathbb{Q}$ defined by $f(x) = x$ is a homomorphism that is one–one but not onto.

(c) The function $f\colon \mathbb{Z} \to \mathbb{Z}_n$ defined by $f(x) = x \pmod n$ is a ring homomorphism that is onto but not one–one.

(d) Let $f\colon \mathbb{C} \to \mathbb{C}$ be the conjugation function $f(a+ib) = a - ib$. Then f is a homomorphism that is one–one and onto.

Exercise 4.3

Show that each of the following functions is a homomorphism.

(a) The function $f\colon \mathbb{Z}[x] \to \mathbb{Z}_2$ defined by $f(r) = r(0) \pmod 2$.

(b) The function $f\colon \mathbb{Q}[\sqrt{5}] \to \mathbb{Q}[\sqrt{5}]$ defined by $f(a+b\sqrt{5}) = a - b\sqrt{5}$.

(c) The function $f\colon \mathbb{Z} \to \mathbb{C}$ defined by $f(x) = x + 0i$.

Since a homomorphism preserves the behaviour of the ring operation, it is not unreasonable to expect some of the ring properties to be preserved too. This is indeed the case.

Proposition 4.9 *Basic properties of homomorphisms*

Let $f: R \to S$ be a ring homomorphism. Then:

(a) $f(0_R) = 0_S$, where 0_R and 0_S are the additive identities of R and S respectively

(b) $f(-a) = -f(a)$ for all $a \in R$

(c) if u is a unit in R, then $f(u)$ is a unit in S.

The proof of Proposition 4.9 is left as an exercise.

Exercise 4.4

Prove Proposition 4.9.

In Example 4.8 we saw two ring homomorphisms that were both one–one and onto. Homomorphisms with these properties are known as isomorphisms, and we already defined them in Definition 1.4. An isomorphism between a ring and itself, such as complex conjugation in Example 4.8(d), has a special name.

Definition 4.10 *Ring automorphism*

A ring isomorphism $f: R \to R$ whose domain and codomain are the same ring R is known as a ring **automorphism**.

As in group theory and linear algebra, there are two special sets associated with a ring homomorphism.

Definition 4.11 *Image and kernel*

Let R and S be rings, and $f: R \to S$ a ring homomorphism.

(a) The **image** of f is the set $\mathrm{Im}(f) = \{f(a) : a \in R\}$.

(b) The **kernel** of f is the set $\mathrm{Ker}(f) = \{a \in R : f(a) = 0\}$.

As in the case of group homomorphisms, the image of a ring homomorphism is a subset of the codomain and the kernel is a subset of the domain, so if $f: R \to S$ is a ring homomorphism then

$$\mathrm{Im}(f) \subseteq S \quad \text{and} \quad \mathrm{Ker}(f) \subseteq R.$$

The kernel and image can be used to check whether or not a given homomorphism is one–one or onto.

Proposition 4.12

Let $f: R \to S$ be a ring homomorphism. Then:

(a) f is onto if, and only if, $\mathrm{Im}(f) = S$

(b) f is one–one if, and only if, $\mathrm{Ker}(f) = \{0\}$.

Proof

(a) If $S = \mathrm{Im}(f)$, then any element of S can be written as $f(r)$ for some $r \in R$, so f is onto.

Conversely, if f is onto then for any $s \in S$ there is $r \in R$ such that $s = f(r)$, so $s \in \mathrm{Im}(f)$. Therefore $S = \mathrm{Im}(f)$.

(b) Suppose f is one–one. If $a \in \mathrm{Ker}(f)$, then $f(a) = 0$, so $f(0) = 0 = f(a)$. But f is one–one, so $a = 0$ and hence $\mathrm{Ker}(f) = \{0\}$.

Conversely, suppose that $\mathrm{Ker}(f) = \{0\}$. If $f(a) = f(b)$, then $f(a - b) = f(a) - f(b) = 0$. Thus $a - b \in \mathrm{Ker}(f) = \{0\}$, and hence $a - b = 0$, so $a = b$. It follows that f is one–one. ■

Example 4.13 *Kernel and image of a homomorphism*

Consider the homomorphism $f: \mathbb{Z}[x] \to \mathbb{Z}$ defined by $f(r) = r(0)$. That is, f maps each polynomial in $\mathbb{Z}[x]$ to its constant coefficient.

We have that $\mathrm{Im}(f)$ is the whole of $\mathbb{Z}$, for if $a \in \mathbb{Z}$ then $f(a + x) = a$, and so $\mathrm{Im}(f) = \mathbb{Z}$.

To find $\mathrm{Ker}(f)$, let $r \in \mathbb{Z}[x]$ with $f(r) = 0$, so $r(0) = 0$. Then

$$\mathrm{Ker}(f) = \{r : r \text{ has constant coefficient } 0\},$$

and $\mathrm{Ker}(f) \neq \{0\}$. It follows that f is onto but not one–one.

Exercise 4.5

Find the image and kernel of each of the following homomorphisms from Exercise 4.3. Decide whether the homomorphism is onto and/or one–one.

(a) $f: \mathbb{Z}[x] \to \mathbb{Z}_2$ defined by $f(r) = r(0) \pmod 2$.

(b) $f: \mathbb{Z}[\sqrt{5}] \to \mathbb{Z}[\sqrt{5}]$ defined by $f(a + b\sqrt{5}) = a - b\sqrt{5}$.

(c) $f: \mathbb{Z} \to \mathbb{C}$ defined by $f(x) = x + 0i$.

You may have noticed that the kernel of f in the solution to Exercise 4.5(a) is the ideal of $\mathbb{Z}[x]$ that we discussed in Worked Exercise 2.12. Again, in each of parts (b) and (c), the kernel of f is the trivial ideal, that is, $\{0\}$. It turns out that the kernel of a homomorphism is always an ideal, and the image is always a subring.

Proposition 4.14

Let R and S be rings, and $f: R \to S$ a ring homomorphism. Then:

(a) $\mathrm{Im}(f)$ is a subring of S

(b) $\mathrm{Ker}(f)$ is an ideal of R.

Proof

(a) Since $0 = f(0)$, $\mathrm{Im}(f)$ contains 0, and so it is non-empty. We check the subring criteria (Book C, Chapter 11, Lemma 1.8). Let $a, b \in \mathrm{Im}(f)$. Then $a = f(c)$ and $b = f(d)$ for some $c, d \in R$, so

SR1 $a - b = f(c) - f(d) = f(c - d) \in \mathrm{Im}(f)$

SR2 $ab = f(c)f(d) = f(cd) \in \mathrm{Im}(f)$

SR3 $f(1) = 1$, so $\mathrm{Im}(f)$ contains the multiplicative identity 1.

It follows that $\mathrm{Im}(f)$ is a subring of S.

(b) Since $\mathrm{Ker}(f)$ contains zero, the additive identity, it must be non-empty. Moreover,

I1 if $a, b \in \mathrm{Ker}(f)$, then $f(a) = f(b) = 0$ and so $f(a - b) = f(a) - f(b) = 0$, giving $a - b \in \mathrm{Ker}(f)$;

I2 if $a \in \mathrm{Ker}(f)$ and $r \in R$, then $f(a) = 0$ and so $f(ra) = f(r)f(a) = 0$, so $ra \in I$.

So the ideal conditions are satisfied, and $\mathrm{Ker}(f)$ is an ideal of R. ■

We can say more, as the following proposition shows.

Proposition 4.15

Let I be an ideal of a ring R. Then there is a homomorphism $f: R \to R/I$ such that $\mathrm{Ker}(f) = I$.

Proof Let f be defined by

$$f(a) = a + I.$$

Then $f(1) = 1 + I$, and for all $a, b \in R$

$$f(a + b) = (a + b) + I = (a + I) + (b + I) = f(a) + f(b), \quad \text{and}$$

$$f(ab) = ab + I = (a + I)(b + I) = f(a)f(b).$$

So f is a homomorphism.

By Proposition 4.2, $f(a) = a + I$ is equal to the additive identity I of R/I if, and only if, $a \in I$. So $\mathrm{Ker}(f) = I$. ■

The function defined in the above proposition is known as the *natural homomorphism* from R to R/I.

Definition 4.16 *Natural homomorphism*

Let I be an ideal of a ring R. Then the **natural homomorphism** from R to R/I is the function

$$\begin{aligned} f \;:\; & R \to R/I \\ & a \mapsto a + I. \end{aligned}$$

Another homomorphism that we will find useful is the modulo function

$$\hat{f}_p : \mathbb{Z}[x] \to \mathbb{Z}_p[x]$$

that takes each polynomial with integer coefficients, and reduces its coefficients modulo some prime number p, that is,

$$\hat{f}_p(a_0 + a_1 x + \cdots + a_n x^n) = \hat{a}_0 + \hat{a}_1 x + \cdots + \hat{a}_n x^n,$$

where $a_i \equiv \hat{a}_i \pmod{p}$, for $i = 1, 2, \ldots, n$. Some texts refer to this as the **canonical homomorphism**. In Book C, Chapter 11, Lemma 4.19, we showed that a polynomial with leading coefficient not divisible by p is irreducible if it is irreducible in $\mathbb{Z}_p$. The proof of the lemma uses the homomorphism $\hat{f}_p$ in an essential way.

Exercise 4.6

Find the kernel of the modulo function $\hat{f}_p$ defined above.

4.3 The First Isomorphism Theorem

If we are given a one–one ring homomorphism $f\colon R \to S$, we cannot in general say that R is isomorphic to S, because f may not be onto. We can, however, say that R is isomorphic to $\mathrm{Im}(f)$. Indeed, we could simply redefine the codomain of f to be $\mathrm{Im}(f)$, thereby turning f into an isomorphism.

But what happens if f is not one–one? We will show that $\mathrm{Im}(f)$ is then isomorphic to the quotient $R/\operatorname{Ker}(f)$. The quotient is a ring because $\operatorname{Ker}(f)$ is an ideal of R by Proposition 4.14(b).

Theorem 4.17 *First Isomorphism Theorem*

Let $f\colon R \to S$ be a ring homomorphism from a ring R to a ring S. Then the function $\psi : R/\operatorname{Ker}(f) \to \mathrm{Im}(f)$ defined by

$$\psi : a + \operatorname{Ker}(f) \mapsto f(a)$$

is an isomorphism.

Proof Firstly, since ψ is defined in terms of a representative a of the coset $a + \operatorname{Ker}(f)$ rather than in terms of $a + \operatorname{Ker}(f)$ itself, we must check that ψ is well defined.

For this, we need to show that if a and $b \in R$ are such that $a + \mathrm{Ker}(f) = b + \mathrm{Ker}(f)$, then $\psi(a + \mathrm{Ker}(f)) = \psi(b + \mathrm{Ker}(f))$. But if $a + \mathrm{Ker}(f) = b + \mathrm{Ker}(f)$ then $(b - a) \in \mathrm{Ker}(f)$ by Proposition 4.2, and so $0 = f(b - a) = f(b) - f(a)$. Hence $f(a) = f(b)$, that is $\psi(a + \mathrm{Ker}(f)) = \psi(b + \mathrm{Ker}(f))$, as required.

Secondly, we check that ψ is a bijection. For this, suppose $\psi(a + \mathrm{Ker}(f)) = 0$. Then $f(a) = 0$, and hence $a \in \mathrm{Ker}(f)$. So $a + \mathrm{Ker}(f) = \mathrm{Ker}(f)$ is the additive identity of $R/\mathrm{Ker}(f)$, and so ψ is one–one by Proposition 4.12(b).

To see that ψ is onto, it suffices to note that an element $f(a) \in \mathrm{Im}(f)$ is the image of $a + \mathrm{Ker}(f) \in R/\mathrm{Ker}(f)$ under ψ.

Finally, we check that ψ satisfies the conditions in Definition 1.4. Let $a + \mathrm{Ker}(f)$ and $b + \mathrm{Ker}(f)$ be arbitrary elements of $R/\mathrm{Ker}(f)$. Then

- $$\begin{aligned}\psi\left((a + \mathrm{Ker}(f)) + (b + \mathrm{Ker}(f))\right) &= \psi\left((a + b) + \mathrm{Ker}(f)\right)\\ &= f(a + b)\\ &= f(a) + f(b)\\ &= \psi\left(a + \mathrm{Ker}(f)\right) + \psi(b + \mathrm{Ker}(f))\end{aligned}$$
- $$\begin{aligned}\psi\left((a + \mathrm{Ker}(f)) \cdot (b + \mathrm{Ker}(f))\right) &= \psi\left(ab + \mathrm{Ker}(f)\right)\\ &= f(ab)\\ &= f(a)f(b)\\ &= \psi\left(a + \mathrm{Ker}(f)\right)\psi\left(b + \mathrm{Ker}(f)\right).\end{aligned}$$

Moreover,

- $\psi(1 + \mathrm{Ker}(f)) = f(1) = 1$, so ψ maps the multiplicative identity of $R/\mathrm{Ker}(f)$ to the multiplicative identity of $\mathrm{Im}(f)$.

Thus, ψ is an isomorphism between $R/\mathrm{Ker}(f)$ and $\mathrm{Im}(f)$, as required. ■

Example 4.18 *The modular ring $\mathbb{Z}_5$*

The isomorphism $\mathbb{Z}/5\mathbb{Z} \cong \mathbb{Z}_5$ can be written in the form

$$\mathbb{Z}/\mathrm{Ker}(f) \cong \mathrm{Im}(f),$$

where $f : \mathbb{Z} \to \mathbb{Z}_5$ is the function defined by

$$f(x) = x \pmod 5.$$

This is because $f(x) = 0 \in \mathbb{Z}_5$ if, and only if, $x \equiv 0 \pmod 5$, that is, x is a multiple of 5, so $\mathrm{Ker}(f) = 5\mathbb{Z}$. Moreover, $\mathrm{Im}(f) = \mathbb{Z}_5$ because of the way that f has been defined.

Hence this isomorphism can be obtained by Theorem 4.17.

In Worked Exercise 4.6 we made the observation that if $I = \langle x^2 + 1\rangle$, then the quotient ring $\mathbb{Q}[x]/I$ looks like the ring $\mathbb{Q}[i]$. In support of this we noted that the ring operations for adding and multiplying cosets are

$$(a + bx + I) + (c + dx + I) = (a + c) + (b + d)x + I,$$
$$(a + bx + I) \cdot (c + dx + I) = (ac - bd) + (ad + bc)x + I.$$

If we ignore every '$+I$' and replace every x by i, we obtain the formulas for adding and multiplying the elements of $\mathbb{Q}[i]$, namely

$$(a+bi)+(c+di)=(a+c)+(b+d)i,$$
$$(a+bi)\cdot(c+di)=(ac-bd)+(ad+bc)i.$$

This suggests that $\mathbb{Q}[x]/I$ is isomorphic to $\mathbb{Q}[i]$. This is indeed the case, and in the next subsection we will use the First Isomorphism Theorem to make these ideas more precise, not just for the ring $\mathbb{Q}[i]$, but also for rings like $\mathbb{Q}[\sqrt{2}]$, $\mathbb{Q}[\sqrt{-3}]$ and $\mathbb{Q}[\sqrt{-5}]$. In Chapter 18 we will see that these are all, in fact, fields.

4.4 Quotient rings of $K[x]$

In this subsection, we will look at quotient rings of $K[x]$, where K is a field and the ideal by which we quotient is generated by an irreducible polynomial. We start with the special case $K=\mathbb{Q}$.

So suppose that $p(x)$ be an irreducible polynomial in $\mathbb{Q}[x]$, and consider the ideal $P=\langle p(x)\rangle$.

In order to relate quotient rings such as $\mathbb{Q}[x]/P$ to rings such as $\mathbb{Q}[i]$, we would like to find a homomorphism from $\mathbb{Q}[x]$ to $\mathbb{C}$, the field of complex numbers, whose kernel is P. This would allow us to use the First Isomorphism Theorem to describe $\mathbb{Q}[x]/P$ as a subring of $\mathbb{C}$.

For a start, we know that the kernel of such a homomorphism is P if any multiple of the irreducible polynomial $p(x)$ is mapped to zero. It will help to introduce certain maps from $\mathbb{Q}[x]$ to $\mathbb{C}$.

Definition 4.19 *Evaluation homomorphism*

Let $\alpha\in\mathbb{C}$. Then the function $f_\alpha:\mathbb{Q}[x]\to\mathbb{C}$ defined by

$$f_\alpha(a_0+a_1x+\cdots+a_nx^n)=a_0+a_1\alpha+\cdots+a_n\alpha^n$$

is called an **evaluation homomorphism**.

The functions f_α defined in this way are indeed homomorphisms, and before we proceed we need to check this. In the next exercise you are asked to show that an evaluation homomorphism satisfies the first condition in Definition 4.7.

Exercise 4.7

Let $\alpha\in\mathbb{C}$, and let f_α be the evaluation homomorphism in Definition 4.19. Show that for p and $q\in\mathbb{Q}[x]$

$$f_\alpha(p+q)=f_\alpha(p)+f_\alpha(q).$$

We now need to check that an evaluation homomorphism f_α satisfies the second condition in Definition 4.7, that is, for p and $q \in \mathbb{Q}[x]$

$$f_\alpha(pq) = f_\alpha(p)f_\alpha(q).$$

Now $f_\alpha(x^m x^n) = f_\alpha(x^{m+n}) = \alpha^{m+n} = \alpha^m\alpha^n = f_\alpha(x^m)f_\alpha(x^n)$. Thus the condition is satisfied by powers of x. It is also the case that

$$f_\alpha(ax^m) = a\alpha^m = f_\alpha(a)f_\alpha(x^m).$$

Putting this together with the solution to Exercise 4.7 shows that the required condition is satisfied in general.

Finally, we note that $f_\alpha(1) = 1$. Thus f_α is a homomorphism.

We are now in a position to show that, given an ideal $P = \langle p(x)\rangle$ in $\mathbb{Q}[x]$, we can find $\alpha \in \mathbb{C}$ such that P is equal to $\mathrm{Ker}(f_\alpha)$.

We mentioned the Fundamental Theorem of Algebra in Book C, Chapter 11, Section 4.1 (see the comments following Exercise 4.1). This theorem tells us that for any $p \in \mathbb{Q}[x]$ there is $\alpha \in \mathbb{C}$ such that α is a root of the polynomial equation $p(x) = 0$, that is

$$p(\alpha) = 0.$$

Since any element of P has the form pq for some $q \in \mathbb{Q}[x]$, and since

$$f_\alpha(pq) = p(\alpha)q(\alpha) = 0,$$

this means that $P \subseteq \mathrm{Ker}(f_\alpha)$. Our aim is to show that this inclusion is in fact an equality. For this, it will help to consider certain special ideals in the more general context of PIDs.

Definition 4.20 *Maximal ideal*

A proper ideal I of a ring R is a **maximal ideal** if whenever J is an ideal of R such that $I \subseteq J \subseteq R$, then either $I = J$ or $J = R$.

For example, $6\mathbb{Z}$ is not a maximal ideal of $\mathbb{Z}$ because it is a proper subset of the proper ideal $3\mathbb{Z}$. By contrast, if $p \in \mathbb{Z}$ is prime, then $p\mathbb{Z}$ is maximal. This is because by Lemma 2.7 any ideal I that contains $p\mathbb{Z}$ must be generated by a divisor of p, so either $I = p\mathbb{Z}$ or $I = 1\mathbb{Z} = \mathbb{Z}$.

Worked Exercise 4.21

Decide whether each of the following ideals of $\mathbb{Q}[x]$ is maximal.

(a) $\langle x^5 + x^4 + x^2 - 1\rangle$

(b) $\langle x^5 - 6x^2 + 9x + 6\rangle$

Solution

(a) The generator $x^5 + x^4 + x^2 - 1$ has -1 as one of its roots, and therefore has $x + 1$ as a proper factor. So $\langle x^5 + x^4 + x^2 - 1\rangle$ is not a maximal ideal because it is a proper subset of the ideal $\langle x + 1\rangle \neq \mathbb{Q}[x]$.

(b) Eisenstein's Criterion (Book C, Chapter 11, Theorem 4.15) for $p = 3$ tells us that $q(x) = x^5 - 6x^2 + 9x + 6$ is irreducible. Suppose that $\langle q(x)\rangle \subseteq I$ for some ideal I of $\mathbb{Q}[x]$. Since $\mathbb{Q}[x]$ is a PID, $I = \langle r(x)\rangle$ for some $r(x)$, and because $q(x)$ is irreducible, $r(x)$ is either a unit or an associate of $q(x)$. If $r(x)$ is a unit, then $\langle r(x)\rangle = \mathbb{Q}[x]$ by Exercise 2.4. If $r(x)$ is an associate of $q(x)$, then $\langle r(x)\rangle = \langle q(x)\rangle$ by Exercise 2.9. Therefore $\langle x^5 - 6x^2 + 9x + 6\rangle$ is a maximal ideal.

The above examples suggest that irreducible elements are the only elements capable of generating maximal ideals of a PID. The following theorem confirms that this is indeed the case.

Theorem 4.22

Let R be a PID, and let a be a non-zero, non-unit element of R. Then $\langle a\rangle$ is a maximal ideal of R if, and only if, a is irreducible in R.

Proof Let a be an irreducible element of R, and let $\langle b\rangle$ be an arbitrary ideal of R such that $\langle a\rangle \subset \langle b\rangle$. Then Lemma 2.7 implies that $a = bc$ for some $c \in R$. Since a is irreducible, either b is a unit, and hence by Exercise 2.4, $\langle b\rangle = R$, or c is a unit, and hence a and b are associates and so, by Exercise 2.9, $\langle a\rangle = \langle b\rangle$. So $\langle a\rangle$ is a maximal ideal.

Conversely, suppose $\langle a\rangle$ is maximal, and let $a = bc$. Then $\langle a\rangle \subseteq \langle b\rangle$. Since $\langle a\rangle$ is maximal, either $\langle b\rangle = \langle a\rangle$, in which case c is a unit by Lemma 2.7, or $\langle b\rangle = R$, in which case b is a unit. So a is irreducible. ■

Exercise 4.8

Which ideals are maximal in $\mathbb{Z}$?

We now return to our discussion about an irreducible polynomial $p(x) \in \mathbb{Q}[x]$. We had chosen $\alpha \in \mathbb{C}$ so that $p(\alpha) = 0$ and established that $\langle p\rangle \subseteq \mathrm{Ker}(f_\alpha)$ for the evaluation map f_α. We know that $\mathrm{Ker}(f_\alpha) \neq \mathbb{Q}[x]$. Theorem 4.22 also tells us that $\langle p\rangle$ is a maximal ideal. This means that $\langle p\rangle = \mathrm{Ker}(f_\alpha)$, and so

$$\mathbb{Q}[x]/\langle p\rangle \cong \mathrm{Im}\, f_\alpha.$$

So now we can explain the similarity between $\mathbb{Q}[x]/\langle x^2+1\rangle$ and $\mathbb{Q}[i]$ that we noticed in the comments following Exercise 4.6. We know that x^2+1 is irreducible in $\mathbb{Q}[x]$, and we know that i is a root of x^2+1 in $\mathbb{C}$. By the above discussion, we have that $\langle x^2+1\rangle$ is the kernel of the evaluation homomorphism f_i, and so

$$\mathbb{Q}[x]/\langle x^2+1\rangle \cong \mathrm{Im}(f_i).$$

But $\mathrm{Im}(f_i) = \{p(i) : p(x) \in \mathbb{Q}[x]\}$, and you can check that $\{p(i) : p(x) \in \mathbb{Q}[x]\} = \mathbb{Q}[i]$. Then

$$\mathbb{Q}[x]/\langle x^2+1\rangle \cong \mathbb{Q}[i]$$

so the 'similarity' is in fact an isomorphism.

The next theorem shows that the quotient rings formed from maximal ideals – that is, those generated by irreducible elements – are the only ones that are fields. It is worth noting that this result holds in arbitrary (commutative) rings, not just in PIDs.

Theorem 4.23

Let I be an ideal of a ring R. Then I is a maximal ideal if, and only if, R/I is a field.

Proof Suppose R/I is a field. Let J be an ideal of R containing I as a proper subset. Then there exists $a \in J$ such that $a \notin I$, and so $a+I$ is a non-zero element of R/I. Since R/I is a field there must be an element $b+I \in R/I$ such that $(a+I)(b+I) = ab+I = 1+I$. Then $1 \in ab+I \subseteq J$, and hence $J = R$. So I is a maximal ideal.

Conversely, suppose I is a maximal ideal of R. To show that R/I is a field we need to check that each of its non-zero elements has an inverse. Note that by 'non-zero' we mean 'different from the additive identity in R/I', that is, different from I. So let $a+I$ be a non-zero element of R/I.

Then $a \notin I$ so the ideal $J = I + \langle a\rangle = \{k+ar : k \in I, r \in R\}$ contains I as a proper subset. Since I is maximal, $J = R$, so $1 \in J$ and hence $1 = k+ar$ for some $r \in R$. Then $1-ar = k \in I$, so $ar+I = 1+I$, that is, $(a+I)(r+I) = 1+I$. It follows that $a+I$ has inverse $r+I$, and hence that R/I is a field. ■

Worked Exercise 4.24

Let $R = \mathbb{Z}_3[x]$, the ring of polynomials over $\mathbb{Z}_3$.

(a) Show that the polynomial x^2+1 is irreducible in R.

(b) Describe the quotient ring $R/\langle x^2+1\rangle$.

Solution

(a) Since $-1 = 2 \pmod 3$, there is no $x \in \mathbb{Z}_3$ such that $x^2 = -1$, that is, $x^2 + 1$ has no root in $\mathbb{Z}_3$. By the Factor Theorem (Book C, Chapter 11, Theorem 4.2), this implies that $x^2 + 1$ is irreducible in $\mathbb{Z}_3$.

(b) Since $x^2 + 1$ is irreducible and R is a Euclidean domain, and so a PID, $\langle x^2 + 1 \rangle$ is a maximal ideal. This means $R/\langle x^2 + 1 \rangle$ is a field.

Exercise 4.9

For each of the following values of $p(x)$, determine whether $p(x)$ is irreducible in $\mathbb{Z}_2[x]$. Hence determine whether $\mathbb{Z}_2[x]/\langle p(x) \rangle$ is a field.

(a) $p(x) = x^2 + 1$

(b) $p(x) = x^2 + x + 1$

4.5 Prime ideals

In the previous subsection we have seen that if M is a maximal ideal in a ring R then the quotient ring R/M is a field, and that in a PID irreducible elements generate maximal ideals. Since a PID is an integral domain, and since we know from Book C, Chapter 12, Theorem 2.12 that in an integral domain prime elements are irreducible, we now know that in a PID prime elements generate maximal ideals.

However, there are rings that are integral domains and not PIDs. A familiar example is the ring of polynomials over the integers, $\mathbb{Z}[x]$. Can we say anything about ideals generated by a prime element in such rings? It turns out that we can, but in order to do so we need to introduce the notion of a *prime ideal*.

Definition 4.25 *Prime ideal*

A proper ideal, P, in a ring R is **prime** if whenever $a, b \in R$ are such that $ab \in P$, then either $a \in P$ or $b \in P$.

In the next exercise you are asked to show that an ideal generated by a prime element is prime, and then to identify the prime ideals in $\mathbb{Z}$.

Exercise 4.10

(a) Let R be a ring. Show that, if $p \in R$ is a prime element, then the ideal $\langle p \rangle$ in R is a prime ideal.

(b) What are the prime ideals in $\mathbb{Z}$?

Just as quotient rings of maximal ideals are special, so are quotient rings of prime ideals.

Theorem 4.26

Let R be a commutative ring and P be an ideal of R. Then P is a prime ideal if, and only if, R/P is an integral domain.

Proof Suppose that P is a prime ideal of R and let $a + P, b + P \in R/P$ be such that

$$(a + P)(b + P) = P, \text{ the zero element of } R/P.$$

Then $ab + P = P$, so $ab \in P$. But P is prime, hence either $a \in P$ or $b \in P$. It follows that either $a + P = P$ or $b + P = P$, and P is the zero element of R/P. Therefore R/P is an integral domain.

Conversely, suppose that R/P is an integral domain and let $a, b \in R$ be such that $ab \in P$. Then $ab + P = P$, the zero element of R/P. Since R/P is an integral domain, either $a + P = P$ or $b + P = P$, that is, either $a \in P$ or $b \in P$. Therefore P is a prime ideal, as required. ■

In the next exercise, you are asked to show that in any ring maximal ideals are prime, but that the converse is not always true.

Exercise 4.11

(a) Let R be a ring and let I be a maximal ideal of R. Show that I is a prime ideal of R.

(b) Give an example of an ideal of $\mathbb{Z}[x]$ that is prime but not maximal.

Hint: for part (b), consider a prime ideal in $\mathbb{Z}$.

Under the stronger assumption that the ring is a PID, the converse to the statement in Exercise 4.11(a) is true.

Proposition 4.27

Let R be a PID and let P be a proper non-trivial ideal. If P is prime then:

(a) $P = \langle p \rangle$ for some prime element $p \in P$

(b) P is maximal in R.

Proof

(a) Since R is a PID, $P = \langle p \rangle$ for some $p \in P$. We need to show that p is a prime element of R. Suppose that $p \mid ab$ for some $a, b \in R$. Then $ab \in P$ by Lemma 2.7. But P is a prime ideal, so either $a \in P$ or $b \in P$. Again by Lemma 2.7, either $p \mid a$ or $p \mid b$. This shows that p is prime in R.

(b) From part (a), $P = \langle p \rangle$ for a prime element p of R. Since R is an integral domain, p is irreducible in R by Theorem 2.12 of Book C, Chapter 12. Then P is maximal by Theorem 4.22. ■

If the ring is a UFD, prime ideals are not necessarily maximal, but we can still say something useful.

Lemma 4.28

Let P be a prime ideal in a ring R. Then if $a_1 a_2 \cdots a_n \in P$ there is an $i \in \{1, \ldots, n\}$ with $a_i \in P$.

Proof We use induction on n. The result is true for $n = 2$ by the definition of a prime ideal. Suppose that it is true for $n = k$ and suppose that $a_1 \cdots a_k a_{k+1} \in P$. Then let $b = a_1 \cdots a_k$ so that $b a_{k+1} \in P$. By the definition of a prime ideal, either $b \in P$ or $a_{k+1} \in P$. If $b \in P$, by the induction hypothesis there is an $a_i \in P$ for $i \in \{1, \ldots, k\}$. Thus there is an $i \in \{1, \ldots, k+1\}$ with $a_i \in P$. The result follows by induction. ■

Proposition 4.29

Let R be a UFD and suppose that P is a non-zero prime ideal in R. Then P contains an irreducible element q.

Proof Let $a \in P$. Since R is a UFD we can write $a = u p_1 p_2 \cdots p_n$ where u is a unit and p_i is an irreducible element of R for $i = 1, \ldots, n$. Then $p_1 p_2 \cdots p_n \in P$. Since P is a prime ideal, Lemma 4.28 gives an $i \in \{1, \ldots, n\}$ such that $p_i \in P$. ■

In the next exercise you are asked to prove a property of prime ideals in a UFD where every prime ideal is maximal.

Exercise 4.12

Let R be a UFD in which every prime ideal is maximal. Let P be a prime ideal in R. Show that $P = \langle q \rangle$ for some irreducible element q in R.

Identifying prime ideals in a commutative ring is often a first step in understanding the ring. In the last exercise of this chapter you are asked to do this for two familiar polynomial rings.

We know from Theorem 3.1 of Book C, Chapter 11 that if K is a field then the polynomial ring $K[x]$ is a Euclidean domain, and hence a PID. Therefore we can apply Proposition 4.27 and conclude that all non-zero prime ideals in $K[x]$ are of the form $\langle p(x) \rangle$, where $p(x)$ is an irreducible polynomial. Furthermore, these ideals are maximal.

Exercise 4.13

Describe the non-zero prime ideals in

(a) $\mathbb{C}[x]$

(b) $\mathbb{R}[x]$.

Learning outcomes

We list those topics on which we may set assessment questions to test your understanding of the chapter.

After working through this chapter you should be able to:

(a) Understand how to find the field of fractions of an integral domain.

(b) Decide whether a given set is an ideal of a ring.

(c) In simple cases show that a given ideal is not a principal ideal.

(d) Find the highest common factor of elements in a Euclidean domain, using the Euclidean Algorithm.

(e) In simple cases determine whether or not a ring is a PID.

(f) Carry out calculations involving sums and products of ideals.

(g) Describe the set of cosets of $\mathbb{Q}[x]/I$ for a given principal ideal I.

(h) Identify the image and kernel of a homomorphism.

(i) Determine whether a given ideal is a maximal ideal.

(j) Determine whether a given ideal is a prime ideal.

Solutions to exercises

Solution to Exercise 1.1

Since multiplication in D is commutative, for any a and b in D we have that $ab = ba$, and so $(a, b) \sim (a, b)$.

The commutativity of multiplication in D also implies that if $ad = bc$ then $cb = da$. This shows that if $(a, b) \sim (c, d)$ then $(c, d) \sim (a, b)$, that is, $\sim$ is symmetric.

Solution to Exercise 1.2

We have to show that if $[a, b] = [a_1, b_1]$ and $[c, d] = [c_1 d_1]$, then $[ac, bd] = [a_1c_1, b_1d_1]$, that is,

$$(ac, bd) \sim (a_1c_1, b_1d_1).$$

So we need to show that

$$(ac)(b_1d_1) = (bd)(a_1c_1).$$

This follows directly from equation $(*)$.

Solution to Exercise 1.3

(a) The common denominator of the coefficients of the polynomial on the left-hand side is 30, so

$$\frac{1}{2} + \frac{2}{3}x + \frac{1}{5}x^3 = \frac{1}{30}\left(15 + 20x + 6x^3\right).$$

(b) Let $q(x) = q_0 + q_1x + \cdots + q_nx^n$. Then for each i from $0, \ldots, n$ we can write $q_i = \frac{c_i}{d_i}$ where $c_i, d_i \in R$ and $d_i \neq 0$. Let d be the product $d_0d_1 \cdots d_n$. Then $p_i = q_id \in R$. Thus $dq(x) = p(x) \in R[x]$ and so $q(x) = \frac{1}{d}p(x)$. The result follows with $f = \frac{1}{d}$.

Solution to Exercise 1.4

By Proposition 4.7 of Book C, Chapter 12, $\mathrm{hcf}(a, b) = h$, where h is the product of those irreducible elements that occur in both the factorisations of a and of b. (We may assume that the irreducible factors are the same by multiplying by suitable units if necessary.)

Then $a = hc$ and $b = hd$, where c and d have no common irreducible factors, and so $\mathrm{hcf}(c, d) = u$, where u is a unit. Moreover, by Lemma 1.2(a) $\frac{a}{b} = \frac{c}{d}$, as required.

Solution to Exercise 1.5

(a) We use induction on n. The case $n = 2$ is proved in the solution to Exercise 4.1 of Book C, Chapter 12. We now assume that the result holds for $n = k$, that is, we assume that $\text{lcm}(b_1, b_2, \ldots, b_k)$ exists. Let $l_k = \text{lcm}(b_1, b_2, \ldots, b_k)$. Then, again by the solution to Exercise 4.1 in Chapter 12, $l_{k+1} = \text{lcm}(l_k, b_{k+1})$ exists. We now show that $l_{k+1} = \text{lcm}(b_1, b_2, \ldots, b_{k+1})$.

(i) $l_{k+1} = l_k r$ and $l_{k+1} = b_{k+1} d_{k+1}$ for some $r, d_{k+1} \in D$. By the induction hypothesis $l = b_i d_i$ for some $d_i \in D, i = 1, \ldots, n$. Thus $l_{k+1} = b_i(d_i r), i = 1, \ldots, n$.

(ii) Suppose that $d = b_i d_i$ for some $d_i, i = 1, \ldots, k+1$. Then by the induction hypothesis $d = ls$ for some $s \in D$ and so by the solution to Exercise 4.1 of Chapter 12, $d = l_{k+1} t$ for some $t \in D$.

We have shown that $l_{k+1} = \text{lcm}(b_1, b_2, \ldots, b_{k+1})$. The result follows by induction.

(b) Let $b = \text{lcm}(b_1, b_2, \ldots, b_n)$, so that $b = b_i c_i$ for some $c_i, i = 1, \ldots, n$. Then

$$\frac{a_i}{b_i} = \frac{a_i c_i}{b_i c_i} = \frac{a_i c_i}{b}$$

as required.

Solution to Exercise 2.1

By I2, $-b = (-1)b \in I$, and so, by I1, $a + b = a - (-b) \in I$.

Solution to Exercise 2.2

The set K is non-empty, since $0 \in K$. Moreover,

I1 if $u, v \in K$, then $(u - v)r = ur - vr = 0 - 0 = 0$, so $u - v \in K$;

I2 if $u \in K$ and $s \in R$, then $ur = 0$ and $(su)r = s(ur) = 0$, so $su \in K$.

Therefore K is an ideal of R.

Solution to Exercise 2.3

We first observe that $0 = q \cdot 0 \in q\mathbb{Q}[x]$, so $q\mathbb{Q}[x]$ is non-empty.
Next we check the ideal conditions.

I1 Let qk_1 and qk_2 be two arbitrary elements of $q\mathbb{Q}[x]$. Then

$$qk_1 - qk_2 = q(k_1 - k_2) \in q\mathbb{Q}[x].$$

I2 Let qk be an arbitrary element of $q\mathbb{Q}[x]$ and let $r \in \mathbb{Q}[x]$. Then

$$r(qk) = q(rk) \in q\mathbb{Q}[x].$$

We conclude that $q\mathbb{Q}[x]$ is an ideal of $\mathbb{Q}[x]$.

Solution to Exercise 2.4

Since u is a unit in R, there is an element $u^{-1} \in R$ such that $uu^{-1} = 1$. Hence $1 \in \langle u \rangle$ and so if $r \in R$, by I2, $r = 1 \cdot r \in \langle u \rangle$. Thus $\langle u \rangle = R$.

Solution to Exercise 2.5

Suppose that F is a field and that I is a non-trivial ideal in F. Let $a \in I$ with $a \neq 0$. Then, since F is a field, $a^{-1} \in F$ and so, by I2, $1 = aa^{-1} \in I$, which implies that $I = F$. Thus F contains no proper non-trivial ideals.

Now suppose that F contains no proper non-trivial ideals. Let $a \in F$ with $a \neq 0$ and consider the ideal $\langle a \rangle$. Since F contains no proper non-trivial ideals, and since $a \neq 0$, we have that $\langle a \rangle = F$ and so $1 \in \langle a \rangle$. This implies there is $r \in F$ with $ar = 1$, that is, $r = a^{-1} \in F$. Hence every non-zero element of F has a multiplicative inverse, and so F is a field as required.

Solution to Exercise 2.6

(a) We check the subring criterion (Book C, Chapter 11, Lemma 1.8).

Since $0 = 0 + 0\alpha$, T contains 0 and so it is non-empty. Let $a + b\alpha, c + d\alpha \in T$, where $a, b, c, d \in R$. Then

SR1 $(a + b\alpha) - (c + d\alpha) = (a - c) + (b - d)\alpha \in T$, because R is a subring of S, and so $a - c$ and $b - d$ are elements of R;

SR2
$$\begin{aligned}(a + b\alpha)(c + d\alpha) &= ac + bd\alpha^2 + (ad + bc)\alpha \\ &= ac + bdr + (ad + bc)\alpha \in T,\end{aligned}$$
since R is a subring of S;

SR3 $1 = 1 + 0\alpha \in T$, so T contains the multiplicative identity 1.

It follows that T is a subring of S.

(b) We need to show that T is the smallest subring of S that contains both R and α. Because of the way it has been defined, T does contain both R and α. Suppose that W is a subring that contains both R and α. If $a, b \in R$ then $a + b\alpha \in W$ by the additive and multiplicative closure of W. Therefore every element of T is an element of W, and so $T \subseteq W$. Hence T is indeed the smallest subring of S that contains both R and α.

Solution to Exercise 2.7

(a) Since $(1 + \sqrt{2})(-1 + \sqrt{2}) = 1$, the element $1 + \sqrt{2}$ is a unit, hence $\langle 1 + \sqrt{2} \rangle = R$ by Exercise 2.4.

(b) Since $1 \neq 2a$ for all $a \in \mathbb{Z}$, $1 \notin \langle 2 \rangle = \{2a\}$. Then $\langle 2 \rangle \neq R$.

Solution to Exercise 2.8

First suppose that $aR \subseteq bR$. Then $a = a \cdot 1 \in aR \subseteq bR$, so $a = rb$ for some $r \in R$. That is, $b \mid a$.

Conversely, suppose $b \mid a$. Then there is an $s \in R$ such that $a = bs$. Now let $ar \in aR$. Then

$$ar = (bs)r = b(sr) \in bR.$$

Therefore $aR \subseteq bR$.

Solution to Exercise 2.9

By Lemma 2.7, $aR = bR$ if, and only if, $b \mid a$ and $a \mid b$.

Now, if $b \mid a$ and $a \mid b$, then a and b are associates by Lemma 2.13(b) of Book C, Chapter 12.

On the other hand, if a and b are associates, then $a = ub$ for some unit u, and $b = u^{-1}a$, so $b \mid a$ and $a \mid b$. It follows that each of aR and bR contain the other, and hence $aR = bR$.

So $aR = bR$ if, and only if, a and b are associates.

Solution to Exercise 2.10

The zero element is a member of I_0 and therefore a member of I, so I is non-empty. Also:

I1 If $a, b \in I$, then $a \in I_i$ and $b \in I_j$ for some integers $i, j \geq 0$, so $a, b \in I_k$ where $k = \max(i, j)$. Since I_k is an ideal, we have $a - b \in I_k \subseteq I$.

I2 If $a \in I$ and $r \in R$, then $a \in I_i$ for some integer $i \geq 0$. Since I_i is an ideal we have $ra \in I_i \subseteq I$.

Hence I is an ideal of R.

Solution to Exercise 2.11

(a) Since $0 \in I_1$ and $0 \in I_2$, the intersection $I_1 \cap I_2$ contains 0, and so it is non-empty. Moreover:

I1 If $a, b \in I_1 \cap I_2$, then $a - b \in I_1$ because I_1 is an ideal, and $a - b \in I_2$ because I_2 is an ideal, so $a - b \in I_1 \cap I_2$.

I2 If $a \in I_1 \cap I_2$ and $r \in R$, then $ar \in I_1$ because I_1 is an ideal of R, and $ar \in I_2$ because I_2 is an ideal of R, so $ar \in I_1 \cap I_2$, as required.

It follows that $I_1 \cap I_2$ is an ideal of R.

(b) Clearly $2 \in 2\mathbb{Z}$ and $7 \in 7\mathbb{Z}$, but $2 - 7 = -5$ is neither a multiple of 2 nor of 7, so $-5 \notin 2\mathbb{Z} \cup 7\mathbb{Z}$. Hence $2\mathbb{Z} \cup 7\mathbb{Z}$ is not closed under subtraction and therefore it is not an ideal of $\mathbb{Z}$.

Solution to Exercise 2.12

A highest common factor exists because $\mathbb{Z}[\sqrt{-2}]$ is a Euclidean domain, and hence a PID, so we can use the Euclidean Algorithm to find the hcf.

We begin by using ordinary complex division to divide the number with larger norm by the number with smaller norm:

Multiply the numerator and the denominator by $(1 - 5\sqrt{-2})$.

$$\frac{9 + 11\sqrt{-2}}{1 + 5\sqrt{-2}} = \frac{119 - 34\sqrt{-2}}{51} = (2 - \sqrt{-2}) + \tfrac{1}{3}(1 + \sqrt{-2}),$$

where $(2 - \sqrt{-2})$ is the element of $\mathbb{Z}[\sqrt{-2}]$ whose Euclidean norm is closest to that of $\frac{1}{51}(119 - 34\sqrt{-2})$. Then multiplying through by $1 + 5\sqrt{-2}$ gives the quotient–remainder form

$$9 + 11\sqrt{-2} = (2 - \sqrt{-2})(1 + 5\sqrt{-2}) + (-3 + 2\sqrt{-2}).$$

Since the remainder $(-3+2\sqrt{-2})$ is still non-zero, we calculate

$$\frac{1+5\sqrt{-2}}{-3+2\sqrt{-2}} = \frac{17-17\sqrt{-2}}{17} = 1-\sqrt{-2} \in \mathbb{Z}[\sqrt{-2}].$$

The remainder is now zero, so the hcf is the previous remainder:

$$\text{hcf}(9+11\sqrt{-2}, 1+5\sqrt{-2}) = -3+2\sqrt{-2}.$$

Solution to Exercise 2.13

(a) You saw in the solution to Exercise 3.1 of Book C, Chapter 12, that $\mathbb{Z}[\sqrt{-2}]$ is a Euclidean domain, so by Theorem 2.11 it must also be a PID.

(b) If $\mathbb{Z}[\sqrt{-5}]$ were a PID then, by Theorem 2.16, it would also be a UFD. However, we know that $\mathbb{Z}[\sqrt{-5}]$ is not a UFD: for example,

$$6 = 3 \cdot 2 = (1+\sqrt{-5}) \cdot (1-\sqrt{-5})$$

exhibits two different factorisations into irreducibles, as shown in the solution to Exercise 2.6(d) of Book C, Chapter 12. Therefore we conclude that $\mathbb{Z}[\sqrt{-5}]$ is not a PID.

Solution to Exercise 3.1

If $a \in A$ then $a = a + 0 \in A + B$, and so $A \subseteq (A+B)$.
Similarly, if $b \in B$, then $b = 0 + b \in A + B$, and so $B \subseteq (A+B)$.

Solution to Exercise 3.2

Let $x \in \langle a \rangle + \langle b \rangle$. Then $x = sa + tb$ for some $s, t \in R$. Since $h \mid a$ and $h \mid b$, we have that $h \mid x$, and hence $x \in \langle h \rangle$ by Lemma 2.7.

Conversely, if $x \in \langle h \rangle$, then $x = rh$ for some $r \in R$. By Theorem 2.13, there exist $s, t \in R$ such that $h = sa + tb$, so

$$x = rh = (rs)a + (rt)b \in \langle a \rangle + \langle b \rangle.$$

It follows that $\langle a \rangle + \langle b \rangle = \langle h \rangle$.

Solution to Exercise 3.3

Since $0 = 0 \cdot 0 \in AB$, AB is non-empty, so we are left to check the ideal conditions.

I1 Let $a_1b_1 + \cdots + a_kb_k$ and
$c_1d_1 + \cdots + c_\ell d_\ell$
be two arbitrary elements of AB. Then their difference is equal to

$$a_1b_1 + \cdots + a_kb_k + (-c_1)d_1 + \cdots + (-c_\ell)d_\ell,$$

which is an element of AB by the way AB has been defined.

I2 Let $a_1b_1 + \cdots + a_kb_k \in AB$ and let $r \in R$. Then

$$r(a_1b_1 + \cdots + a_kb_k) = (ra_1)b_1 + \cdots + (ra_k)b_k \in AB,$$

where I2 for A ensures that each $ra_i \in A$.

Hence AB is an ideal.

Solution to Exercise 3.4

(a) The ideal $\langle 27, 36\rangle$ must be principal because $\mathbb{Z}$ is a PID.

So let h be a generator of $\langle 27, 36\rangle$. Then h is a common factor of every element of $\langle 27, 36\rangle$, and in particular of 27 and 36. But $h \in \langle 27, 36\rangle$ so $h = 27a + 36b$ for some $a, b \in \mathbb{Z}$. Therefore every common factor of 27 and 36 is also a factor of h. In other words, $h = \text{hcf}(27, 36) = 9$.

Alternatively, since $\langle 27, 36\rangle$ is equal to $\langle 27\rangle + \langle 36\rangle$, by the solution to Exercise 3.2 the generator must be equal to $\text{hcf}(27, 36) = 9$.

(b) The ideal $\langle x, 2\rangle$ is not principal. In fact, we claim it is the ideal A of polynomials in $\mathbb{Z}[x]$ with even constant terms, which we showed to be non-principal in the solution to Worked Exercise 2.12. So it is sufficient to prove that $\langle x, 2\rangle = A$.

Since both 2 and x have an even constant term, they are both elements of A. The closure properties of A then ensure that $\langle x, 2\rangle \subseteq A$.

Conversely, every polynomial in A has an even constant term, and therefore has the form

$$2a + x(a_1 + a_2x + \cdots + a_nx^{n-1}) \in \langle x, 2\rangle,$$

so $A \subseteq \langle x, 2\rangle$. Thus $A = \langle x, 2\rangle$, as required.

Solution to Exercise 3.5

(a) Since $\mathbb{Z}[i]$ is a Euclidean domain, A must be a principal ideal generated by $\text{hcf}(3 + i, 1 + 3i)$. In the solution to Exercise 2.14 it was shown that $\text{hcf}(3 + i, 1 + 3i) = (-1 - i)$, so $A = \langle -1 - i\rangle$.

(b) The ring $\mathbb{Z}[\sqrt{-5}]$ is not a PID, so A might not be principal. We will use the multiplicative norm $N(a + b\sqrt{-5}) = a^2 + 5b^2$ to prove that it is not principal.

Suppose A is generated by some element $t \in A$. We know that $2 \in A$ and $(1 + \sqrt{-5}) \in A$, so $2 = rt$ and $(1 + \sqrt{-5}) = st$ for some $r, s \in \mathbb{Z}[\sqrt{-5}]$.

Taking norms shows that $N(t)$ must be a factor of both $N(2) = 4$ and $N(1 + \sqrt{-5}) = 6$. So either $N(t) = 2$ or $N(t) = 1$. But $N(u + v\sqrt{-5}) = u^2 + 5v^2 = 2$ has no integer solutions, and the only integer solutions of $u^2 + 5v^2 = 1$ are $u = \pm 1$ with $v = 0$, so $t = \pm 1$. But if this value of t really does lie in A, then there are integers a, b, c, d such that

$$\begin{aligned}\pm 1 &= 2(a + b\sqrt{-5}) + (1 + \sqrt{-5})(c + d\sqrt{-5})\\ &= (2a + c - 5d) + (2b + c + d)\sqrt{-5}.\end{aligned}$$

But this implies $2b + c + d = 0$, and hence $c = -2b - d$. So

$$\pm 1 = 2a + c - 5d = 2a - 2b - 6d.$$

But the right-hand side is even, which is a contradiction.

Therefore the assumption that A is generated by t must be false, and so A is not principal.

Solution to Exercise 3.6

(a) Since $x^2 - 1 = (x+1)(x-1) \in \langle x-1 \rangle$ and $x^2 - 2x + 1 = (x-1)^2 \in \langle x-1 \rangle$, we have $\langle x^2-1, x^2-2x+1 \rangle \subseteq \langle x-1 \rangle$ by Proposition 3.8. In fact, the ideal $\langle x^2-1, x^2-2x+1 \rangle$ is a *proper* subset of the ideal $\langle x-1 \rangle$ but this is a bit tricky to show.

(b) (i) We have $(1+\sqrt{-5}) = 2 - (1-\sqrt{-5})$. Hence

$$1 + \sqrt{-5} \in \langle 2, 1 - \sqrt{-5} \rangle,$$

and so $\langle 2, 1+\sqrt{-5} \rangle \subseteq \langle 2, 1-\sqrt{-5} \rangle$ by Proposition 3.8.

Similarly, since $(1-\sqrt{-5}) = 2 - (1+\sqrt{-5})$ we have that

$$\langle 2, 1-\sqrt{-5} \rangle \subseteq \langle 2, 1+\sqrt{-5} \rangle.$$

Hence $\langle 2, 1+\sqrt{-5} \rangle = \langle 2, 1-\sqrt{-5} \rangle$, as required.

(ii) Since $\langle 2-\sqrt{-5} \rangle$ is closed under multiplication by elements of $\mathbb{Z}[\sqrt{-5}]$, and in particular by the element $2+\sqrt{-5}$, we have

$$(2+\sqrt{-5})(2-\sqrt{-5}) = 9 \in \langle 2-\sqrt{-5} \rangle.$$

Hence $\langle 9 \rangle \subseteq \langle 2-\sqrt{-5} \rangle$ by Proposition 3.8.

Solution to Exercise 3.7

First we note that $\langle I, t \rangle$ is non-empty, since $t \in \langle I, t \rangle$. Moreover:

I1 Let $b_1, b_2 \in R$ and $m_1, m_2 \in I$. Then

$$(m_1 + b_1 t) - (m_2 + b_2 t) = (m_1 - m_2) + (b_1 - b_2)t.$$

Since I is an ideal, $m_1 - m_2 \in I$, and so $(m_1 - m_2) + (b_1 - b_2)t \in \langle I, t \rangle$.

I2 Let $b, r \in R$ and $m \in I$. Then

$$(m + bt)r = mr + btr.$$

Since I is an ideal, $mr \in I$. Since R is commutative, $btr = brt$, and so $(mr + btr) \in \langle I, t \rangle$.

Solution to Exercise 4.1

Let $a+I$, $b+I$, $c+I$ be arbitrary cosets of I by a, b, c, respectively.

The set of cosets with addition, $(S, +)$, is a quotient group of an abelian group, and so it is itself an abelian group. We need to show the remaining axioms of a commutative ring are satisfied. In what follows, we will, as usual, replace $\cdot$ by juxtaposition.

Axioms for multiplication

R6 Closure The closure of $\cdot$ is immediate from the definition because $(ab) + I$ is clearly a coset.

R7 Associativity We have

$$\begin{aligned}(a+I)((b+I)(c+I)) &= a(bc) + I \\ &= (ab)c + I \\ &= ((a+I)(b+I))(c+I).\end{aligned}$$

R8 Multiplicative identity The multiplicative identity is $1 + I$, for

$$(1 + I)(a + I) = 1a + I = a + I.$$

R9 Distributive laws For left distributivity, we have

$$\begin{aligned}(a + I)\left((b + I) + (c + I)\right) &= a(b + c) + I \\ &= (ab + ac) + I \\ &= (ab + I) + (ac + I) \\ &= (a + I)(b + I) + (a + I)(c + I),\end{aligned}$$

so distributivity in S follows from distributivity in R. Right distributivity follows from commutativity in R.

R10 Commutativity

$$\begin{aligned}(a + I)(b + I) &= ab + I \\ &= ba + I \\ &= (b + I)(a + I)\end{aligned}$$

The ring axioms R1–R9 hold for S even when R is not commutative. Only right distributivity needs to be checked in addition to the above, and this is similar to left distributivity.

Solution to Exercise 4.2

(a) Each coset has the form $f(x) + I$, where $f \in \mathbb{Q}[x]$. By the Division Algorithm, we can write

$$f(x) = q(x)(x^2 + x + 1) + r(x),$$

where $q, r \in \mathbb{Q}[x]$ with $\deg(r) < \deg(x^2 + x + 1) = 2$. So

$$f(x) - r(x) = q(x)(x^2 + x + 1) \in I.$$

By Proposition 4.2,

$$f(x) + I = r(x) + I.$$

But $\deg(r) < 2$, so $r(x)$ has the form $a + bx$, and

$$R/I = \{(a + bx) + I : a, b \in \mathbb{Q}\}.$$

(b) If $a + bx + I$ and $c + dx + I$ are arbitrary elements of R/I, then

$$\begin{aligned}(a + bx + I) + (c + dx + I) &= (a + bx) + (c + dx) + I \\ &= (a + c) + (b + d)x + I \quad \text{and} \\ (a + bx + I) \cdot (c + dx + I) &= (a + bx) \cdot (c + dx) + I \\ &= ac + (ad + bc)x + bdx^2 + I.\end{aligned}$$

Now $x^2 + x + 1 \in I$, so $x^2 + I = -x - 1 + I$. Then

$$\begin{aligned}ac + (ad + bc)x + bdx^2 + I &= ac + (ad + bc)x + bd(-x - 1) + I \\ &= (ac - bd) + (ad + bc - bd)x + I.\end{aligned}$$

Thus the formula for multiplication is

$$(a + bx + I)(c + dx + I) = (ac - bd) + (ad + bc - bd)x + I.$$

Solution to Exercise 4.3

(a) Let $r, s \in \mathbb{Z}[x]$. Then $f(1) = 1$,

$$\begin{aligned} f(r+s) &= (r+s)(0) \\ &= r(0) + s(0) \\ &\equiv f(r) + f(s) \pmod 2, \quad \text{and} \\ f(rs) &= rs(0) \\ &= r(0)s(0) \\ &\equiv f(r)f(s) \pmod 2, \end{aligned}$$

so f is a ring homomorphism.

(b) Let $a + b\sqrt{5}, c + d\sqrt{5} \in \mathbb{Q}[\sqrt{5}]$. Then $f(1) = 1$, and

$$\begin{aligned} f\left((a+b\sqrt{5}) + (c+d\sqrt{5})\right) &= f\left((a+c) + (b+d)\sqrt{5}\right) \\ &= (a+c) - (b+d)\sqrt{5} \\ &= (a - b\sqrt{5}) + (c - d\sqrt{5}) \\ &= f\left(a+b\sqrt{5}\right) + f\left(c+d\sqrt{5}\right), \end{aligned}$$

$$\begin{aligned} f\left((a+b\sqrt{5})(c+d\sqrt{5})\right) &= f\left((ac+5bd) + (ad+bc)\sqrt{5}\right) \\ &= (ac+5bd) - (ad+bc)\sqrt{5} \\ &= (a - b\sqrt{5})(c - d\sqrt{5}) \\ &= f\left(a+b\sqrt{5}\right) f\left(c+d\sqrt{5}\right), \end{aligned}$$

so f is a homomorphism.

(c) Let $a, b \in \mathbb{Z}$. Then $f(1) = 1$,

$$f(a+b) = a + b + 0i = (a+0i) + (b+0i) = f(a) + f(b), \text{ and}$$
$$f(ab) = ab + 0i = (a+0i)(b+0i) = f(a)f(b),$$

so f is a homomorphism.

Solution to Exercise 4.4

(a) Since $0_R + 0_R = 0_R$, we have

$$f(0_R) = f(0_R + 0_R) = f(0_R) + f(0_R).$$

Adding $-f(0_R)$ to both sides gives $0_S = f(0_R)$, as required.

(b) Since $a + (-a) = 0_R$, we have

$$0_S = f(0_R) = f(a + (-a)) = f(a) + f(-a).$$

But additive inverses are unique, so $f(-a) = -f(a)$, as required.

(c) If $u \in R$ is a unit, then $ur = 1$ for some $r \in R$. Then

$$1_S = f(1_R) = f(ur) = f(u)f(r).$$

So $f(u)$ is a unit, as required.

Solution to Exercise 4.5

(a) We have $f(0) = 0$ and $f(1) = 1$. So $\mathrm{Im}(f) = \mathbb{Z}_2$, and f is onto.

Now let $r \in \mathbb{Z}[x]$ be such that $f(r) = 0$, that is, such that $r(0) \equiv 0 \pmod 2$. Then the constant term of r is even. This shows that

$$\mathrm{Ker}(f) = \{a_0 + a_1x + \cdots + a_nx^n : a_0 \text{ is even}\},$$

and so f is not one–one.

(b) Let $c \in \mathbb{Z}[\sqrt{5}]$. Then $c = a + b\sqrt{5}$ for some $a, b \in \mathbb{Z}$, and so $c = f\left(a - b\sqrt{5}\right)$. Therefore $\mathrm{Im}(f) = \mathbb{Z}[\sqrt{5}]$, and f is onto.

Now let $f(a + b\sqrt{5}) = a - b\sqrt{5} = 0$. Then $a = b = 0$, and hence $a + b\sqrt{5} = 0$. So $\mathrm{Ker}(f) = \{0\}$ and f is one–one.

(c) Here $\mathrm{Im}(f) \neq \mathbb{C}$ since $i \notin \mathrm{Im}(f)$. So f is not onto. In fact, $\mathrm{Im}(f) = \{a + 0i : a \in \mathbb{Z}\}$.

Now let $0 + 0i = f(a) = a + 0i$. Then $a = 0$ and hence $\mathrm{Ker}(f) = \{0\}$. So f is one–one.

Solution to Exercise 4.6

We have $\hat{f}_p(a_0 + a_1x + \cdots + a_nx^n) = 0 + 0x + \cdots + 0x^n$ if, and only if, $a_i \equiv 0 \pmod p$ for $i = 1, 2, \ldots, n$. The latter condition holds if, and only if, each of the coefficients $a_0, a_1, \ldots, a_n$ is an integer multiple of p. So $\mathrm{Ker}(\hat{f}_p) = p\mathbb{Z}[x]$.

Solution to Exercise 4.7

Let $p(x) = a_0 + a_1x + \cdots + a_nx^n$ and $q(x) = b_0 + b_1x + \cdots + b_mx^m$. We assume without loss of generality that $m \geq n$. Then

$$\begin{aligned} f_\alpha(p + q) &= f_\alpha(a_0 + b_0 + (a_1 + b_1)x + \cdots + (a_n + b_n)x^n + b_{n+1}x^{n+1} + \cdots + b_mx^m) \\ &= a_0 + b_0 + (a_1 + b_1)\alpha + \cdots + (a_n + b_n)\alpha^n + b_{n+1}\alpha^{n+1} + \cdots + b_m\alpha^m \\ &= a_0 + a_1\alpha + \cdots + a_n\alpha^n + b_0 + b_1\alpha + \cdots + b_m\alpha^m \\ &= f_\alpha(p) + f_\alpha(q). \end{aligned}$$

Solution to Exercise 4.8

Since $\mathbb{Z}$ is a PID, we know from Theorem 4.22 that maximal ideals have the form $\langle p \rangle$ where p is an irreducible element, that is, an integer prime.

Solution to Exercise 4.9

(a) Since 1 is a root of $x^2 + 1$ in $\mathbb{Z}_2$, the Factor Theorem tells us that $x^2 + 1$ is reducible. In fact

$$x^2 + 1 \equiv x^2 + 2x + 1 \equiv (x + 1)(x + 1) \pmod 2.$$

Since $\mathbb{Z}_2[x]$ is a PID, $\langle x^2 + 1 \rangle$ is not a maximal ideal by Theorem 4.22, and so $\mathbb{Z}_2[x]/\langle x^2 + 1 \rangle$ is not a field by Theorem 4.23.

(b) If $p(x) = x^2 + x + 1$ is reducible over $\mathbb{Z}_2$, either $p(0) = 0$ or $p(1) = 0$ by the Factor Theorem. But $p(0) = p(1) = 1$, and so $p(x)$ is irreducible over $\mathbb{Z}_2$. Then $\mathbb{Z}_2[x]/\langle x^2 + x + 1 \rangle$ is a field.

Solution to Exercise 4.10

(a) Let $ab \in \langle p \rangle$, so that $ab = pr$ for some $r \in R$. So $p \mid ab$, and since the element p is prime in R, either $p \mid a$ or $p \mid b$. Without loss of generality we assume $p \mid a$, so there is $s \in R$ such that $ps = a$, which implies that $a \in \langle p \rangle$, as required.

(b) Since $\mathbb{Z}$ is a PID all its ideals are of the form $\langle a \rangle$. From part (a) we know that if a is a prime integer then $\langle a \rangle$ is a prime ideal. If a is non-zero and not prime, then $a = bc$ for some $b, c \neq 1$ so that $bc \in \langle a \rangle$ but $b \notin \langle a \rangle$ and $c \notin \langle a \rangle$. Thus the non-zero prime ideals of $\mathbb{Z}$ are all the ideals $\langle a \rangle$ where a is a prime integer. Since $\mathbb{Z}$ has no zero divisors, $\langle 0 \rangle$ is also a prime ideal in $\mathbb{Z}$.

Solution to Exercise 4.11

(a) Since I is maximal, by Theorem 4.23, R/I is a field and hence it is an integral domain. Hence I is prime by Theorem 4.26.

(b) The ideal $\langle 5 \rangle$ is prime in $\mathbb{Z}[x]$ by Lemma 4.10 of Book C, Chapter 12 and by the solution to Exercise 4.10(a). However, $\langle 5 \rangle$ is not maximal, since it is properly contained, say, in $\langle 5, x \rangle$, which in turn is properly contained in $\mathbb{Z}[x]$.

Solution to Exercise 4.12

By Proposition 4.29, since R is a UFD, P contains an irreducible element Q. Since R is a UFD, q is a prime element by Corollary 4.3 of Book C, Chapter 12. Thus $\langle q \rangle$ is a prime ideal by the solution to Exercise 4.10(a), and since $q \in P, \langle q \rangle \subseteq P$. Since every prime ideal is maximal, we must have $\langle q \rangle = P$ as required.

Solution to Exercise 4.13

(a) Irreducible polynomials in $\mathbb{C}[x]$ can all be written in the form $ax - b$, for some $a, b \in \mathbb{C}$ with $a \neq 0$, and so the non-zero prime ideals of $\mathbb{C}[x]$ are all the ideals $\langle x - c \rangle$ for some $c \in \mathbb{C}$ (we can divide $ax - b$ through by a as a is non-zero). The Fundamental Theorem of Algebra and the Factor Theorem tell us that there are no other irreducible polynomials in $\mathbb{C}[x]$, so there will be no other non-zero prime ideals.

(b) As for $\mathbb{C}[x]$, all the ideals of the form $\langle x - c \rangle$ for some $c \in \mathbb{R}$ are prime in $\mathbb{R}[x]$. In addition, ideals of the form

$$\langle x^2 + bx + c \rangle,$$

where $x^2 + bx + c$ is irreducible, are also prime. There are no other irreducible polynomials in $\mathbb{R}[x]$, and so there are no other non-zero prime ideals.